MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Sept. 1998, p. 807–813
1092-2172/98/$04.00+0

Vol. 62, No. 3

Revision of the Nomenclature for the *Bacillus thuringiensis* Pesticidal Crystal Proteins

N. CRICKMORE,[1] D. R. ZEIGLER,[2] J. FEITELSON,[3] E. SCHNEPF,[3] J. VAN RIE,[4] D. LERECLUS,[5] J. BAUM,[6] AND D. H. DEAN[2]*

School of Biological Sciences, University of Sussex, Brighton, United Kingdom[1]; Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210[2]; Mycogen Corp., San Diego California 92121[3]; Plant Genetic Systems, n. v., Ghent, Belgium[4]; Unité de Biochimie Microbienne, Institut Pasteur, Paris, France[5]; and Ecogen, Inc., Langhorne, Pennsylvania 19047[6]

BACKGROUND AND HISTORY OF PESTICIDAL CRYSTAL PROTEIN NOMENCLATURE 807
PROPOSED NOMENCLATURE 807
ROBUSTNESS OF THE NOMENCLATURE 808
ACKNOWLEDGMENTS 811
REFERENCES 811

BACKGROUND AND HISTORY OF PESTICIDAL CRYSTAL PROTEIN NOMENCLATURE

Since the first cloning of an insecticidal crystal protein gene from *Bacillus thuringiensis* (91), many other such genes have been isolated. Initially, each newly characterized gene or protein received an arbitrary designation from its discoverers: *icp* (64); *cry* (21, 121); *kurhd1* (31); Bta (88); bt1, bt2, etc. (40); type B and type C (43); and 4.5 kb, 5.3 kb, and 6.6 kb (55). The first systematic attempt to organize the genetic nomenclature relied on the insecticidal activities of crystal proteins for the primary ranking of their corresponding genes (44). The *cryI* genes encoded proteins toxic to lepidopterans; *cryII* genes encoded proteins toxic to both lepidopterans and dipterans; *cryIII* genes encoded proteins toxic to coleopterans; and *cryIV* genes encoded proteins toxic to dipterans alone.

This system provided a useful framework for classifying the ever-expanding set of known genes. Inconsistencies existed in the original scheme, however, due to attempts to accommodate genes that were highly homologous to known genes but did not encode a toxin with a similar insecticidal spectrum. The *cryIIB* gene, for example, received a place in the lepidopteran-dipteran class with *cryIIA*, even though toxicity against dipterans could not be demonstrated for the toxin designated CryIIB. Other anomalies arose after the nomenclature was established. The protein named CryIC, for example, was reported to be toxic to both dipterans and lepidopterans (103), while the protein designated CryIB was reported to be toxic to both lepidopterans and coleopterans (8). Because the nomenclature system provided no central committee or database to maintain standardization, new genes encoding a diverse set of proteins without a common insecticidal activity each received the name *cryV*, based on the next available Roman numeral (32, 46, 67, 100, 102, 108).

PROPOSED NOMENCLATURE

We propose in this review a revised nomenclature for the *cry* and *cyt* genes. To organize the wealth of data produced by genomic sequencing efforts, a new nomenclatural paradigm is emerging, exemplified by the internationally recognized cytochrome P-450 superfamily nomenclature system (68a, 122a). Our proposal conforms closely to this model both in conceptual basis and in nomenclature format. The underlying basis of this type of system is to assign names to members of gene superfamilies according to their degree of evolutionary divergence as estimated by phylogenetic tree algorithms. The nomenclature format in such a system is designed to convey rich informational content about these relationships by appending to the mnemonic root a series of numerals and letters assigned in a hierarchical fashion to indicate degrees of phylogenetic divergence. This change from a function-based to a sequence-based nomenclature allows closely related toxins to be ranked together and removes the necessity for researchers to bioassay each new protein against a growing series of organisms before assigning it a name.

In our proposed revision, Roman numerals have been exchanged for Arabic numerals in the primary rank (e.g., Cry1Aa) to better accommodate the large number of expected new proteins. The mnemonic Cyt to designate crystal proteins showing a general cytolytic activity in vitro has been retained because of its historical precedent and entrenchment in the research literature. Our definition of a Cry protein is rather broad: a parasporal inclusion (crystal) protein from *B. thuringiensis* that exhibits some experimentally verifiable toxic effect to a target organism, or any protein that has obvious sequence similarity to a known Cry protein. Similarly, Cyt denotes a parasporal inclusion (crystal) protein from *B. thuringiensis* that exhibits hemolytic activity, or any protein that has obvious sequence similarity to a known Cyt protein. By these criteria, the nontoxic 40-kDa crystal protein from *B. thuringiensis* subsp. *thompsoni*, for example, has been excluded from our list, but the lepidopteran-active 34-kDa protein (now Cry15A) encoded by an adjacent gene has been included (11).

The freely available software applications CLUSTAL W (110) and PHYLIP (27) define the sequence relationships among the toxins to form the framework of the new nomenclature. In the first step, CLUSTAL W aligns the deduced amino acid sequences of the full-length toxins and produces a distance matrix, quantitating the sequence similarities among the set of toxins. CLUSTAL W default settings are employed, except that the "delay divergent sequences" setting in the multiple-alignment parameter menu is reduced from 40 to 0%. The NEIGHBOR application within the PHYLIP package then constructs a phylogenetic tree from the distance matrix by an unweighted pair-group method using arithmetic averages

* Corresponding author. Mailing address: Department of Biochemistry, 484 West Twelfth Ave., Columbus, OH 43210. Phone: (614) 292-8829. Fax: (614) 292-6773. E-mail: dean.10@osu.edu.

(UPGMA) algorithm. The TREEVIEW application (73), with the "phylogenetic tree" and "ladderize left" options selected, produces a graphic presentation of the resulting tree.

We have applied this procedure to the set of holotype sequences given in Table 1 to produce the phylogenetic tree presented in Fig. 1. Vertical lines drawn through the tree show the boundaries used to define the various nomenclatural ranks. The name given to any particular toxin depends on the location of the node where the toxin enters the tree relative to these boundaries. A new toxin that joins the tree to the left of the leftmost boundary will be assigned a new primary rank (an Arabic number). A toxin that enters the tree between the left and central boundaries will be assigned a new secondary rank (an uppercase letter). It will have the same primary rank as the other toxins within that cluster. A toxin that enters the tree between the central and right boundaries will be assigned a new tertiary rank (a lowercase letter). Finally, a toxin that joins the tree to the right of the rightmost boundary will be assigned a new quaternary rank (another Arabic number). Toxins with identical sequences but isolated independently will receive separate quaternary ranks.

By this method each toxin will be assigned a unique name incorporating all four ranks. A completely novel toxin would currently be assigned the name Cry23Aa1. For the sake of convenience, however, we propose that the inclusion of the tertiary rank a and quaternary rank 1 be optional, their use dictated only by a need for clarity. This new toxin could therefore simply be referred to as Cry23A.

In choosing locations for rank boundaries, we attempted to construct a nomenclature reflecting significant evolutionary relationships while at the same time minimizing changes from the gene names assigned under the old system. In the resulting system, proteins with a common primary rank are similar enough that the percent identity can be defined with some confidence. Proteins with the same primary rank often affect the same order of insect; those with different secondary and tertiary ranks may have altered potency and targeting within an order. At the tertiary rank, differences can be due to the accumulation of dispersed point mutations, but often they appear to have resulted from ancestral recombination events between genes differing at a lower rank level (9). The quaternary rank was established to group "alleles" of genes coding for known toxins that differ only slightly, either because of a few mutational changes or an imprecision in sequencing. To avoid confusion, however, the reader should bear in mind the differences between the quaternary rank number and the classical concept of the allele. Any *cry* gene specified with a quaternary rank is a natural isolate. No assumption about functionality is implied by the presence of this rank number in the gene name. In contrast, an allele number would be assumed, unless parenthetical or subscripted information indicated otherwise, to denote a nonfunctional mutant form of a wild-type gene found at a discrete genetic locus. Because of the somewhat modular nature of the Cry proteins and the effect that various segmental relationships could have on the clustering algorithm, it is likely that these boundaries will move slightly or even bend as the addition of new sequences changes the topology of the phylogenetic tree. Currently the boundaries represent approximately 95, 78, and 45% sequence identity.

A *B. thuringiensis* Pesticidal Crystal Protein Nomenclature Committee, consisting of the authors of this paper, will remain as a standing committee of the *Bacillus* Genetic Stock Center (BGSC) to assist workers in the field of *B. thuringiensis* genetics in assigning names to new Cry and Cyt toxins. The corresponding gene or protein sequences must first be deposited into a publicly accessible database (GenBank, EMBL, or PIR) and released by the repository for electronic publication in the database so that the scientific community may conduct an independent analysis. Researchers should submit new sequences directly to the BGSC director (D. R. Zeigler), either by electronic mail (zeigler.1@osu.edu) or on computer diskette. The director will analyze the amino acid sequence as described above and suggest the appropriate name, subject to the approval of the committee. The committee will periodically review the literature of the Cry and Cyt toxins and publish a comprehensive list. This list, alongside other relevant information, will also be available via the Internet at the following URL: http://www.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/.

The current list of *cry* and *cyt* genes (including quaternary ranks) is given in Table 1. New gene names are listed with their previous names, their GenBank accession numbers, and published references. The quaternary ranks were assigned in the order that the gene sequences were discovered in the literature or submitted to the committee. Genes assigned the quaternary rank 1 represent holotype sequences.

The boundaries shown in Fig. 1 allow most *cry* genes to retain the names they received under the system of Höfte and Whiteley (44), after a substitution of Arabic for Roman numerals. There are a few notable exceptions: *cryIG* becomes *cry9A*, *cryIIIC* becomes *cry7Aa*, *cryIIID* becomes *cry3C*, *cryIVC* becomes *cry10A*, *cryIVD* becomes *cry11A*, *cytA* becomes *cyt1A*, and *cytB* becomes *cyt2A* (Table 1). Under the revised system, the known Cry and Cyt proteins fall into 24 sets at the primary rank—Cyt1, Cyt2, and Cry1 through Cry22.

ROBUSTNESS OF THE NOMENCLATURE

The robustness of the current naming process was assessed by a number of additional analyses. The choice of clustering algorithm (unweighted pair-group method using arithmetic averages) was driven largely by the consistent location of a root and constant branch lengths, resulting in a common vertical alignment of sequence names and essentially allowing a "ruler across the tree" approach to naming. It has the drawback of imposing a common evolutionary clock on the clustering process, an assumption that cannot be assured. The distance metric related to percent identity (essentially 1 minus the fraction of identical residues of the total compared without gaps) is the one most commonly found as the output of sequence comparison programs, including CLUSTAL W. For phylogenetic analysis, a more usual distance metric relates to the number of substitutions per site to convert one sequence to the other (e.g., Dayhoff's point accepted mutation [PAM]) and accounts for the possibility of multiple substitutions per site as the sequences are more divergent. The latter method has the drawback of being more computationally intensive, and, for very divergent sequences, requiring too large a value, resulting in numeric computation failures. They also differ in the way sequences of unequal length are handled, with the percent identity method typically ignoring excess sequence and the other methods assigning a penalty. This is particularly important for crystal proteins, since a number of them lack the C-terminal protoxin segments yet are quite related to some longer toxins in the N-terminal toxin segment; we feel that the stronger association of such relationships found by the percent identity method is preferred.

To assess the effect of using the neighbor-joining method to generate an unrooted tree, CLUSTAL W routines were used to generate such a tree with 1,000 bootstraps of the sequence alignment we used for Fig. 1. When an appropriate outgroup was chosen, the resulting tree (not shown) resembled our Fig. 1. The bootstrap values indicated that the tree thus generated

TABLE 1. Known *cry* and *cyt* gene sequences with revised nomenclature assignments

Revised gene name	Original gene or protein name	Accession no.	Coding region[a]	Reference
cry1Aa1	*cryIA(a)*	M11250	527–4054	92
cry1Aa2	*cryIA(a)*	M10917	153–>2955	98
cry1Aa3	*cryIA(a)*	D00348	73–3600	99
cry1Aa4	*cryIA(a)*	X13535	1–3528	62
cry1Aa5	*cryIA(a)*	D17518	81–3608	113
cry1Aa6	*cryIA(a)*	U43605	1–>1860	63
cry1Ab1	*cryIA(b)*	M13898	142–3606	119
cry1Ab2	*cryIA(b)*	M12661	155–3622	111
cry1Ab3	*cryIA(b)*	M15271	156–3620	31
cry1Ab4	*cryIA(b)*	D00117	163–3627	50
cry1Ab5	*cryIA(b)*	X04698	141–3605	40
cry1Ab6	*cryIA(b)*	M37263	73–3537	37
cry1Ab7	*cryIA(b)*	X13233	1–3465	36
cry1Ab8	*cryIA(b)*	M16463	157–3621	69
cry1Ab9	*cryIA(b)*	X54939	73–3537	13
cry1Ab10	*cryIA(b)*	A29125	—[b]	28
cry1Ac1	*cryIA(c)*	M11068	388–3921	3
cry1Ac2	*cryIA(c)*	M35524	239–3769	117
cry1Ac3	*cryIA(c)*	X54159	339–>2192	18
cry1Ac4	*cryIA(c)*	M73249	1–3534	84
cry1Ac5	*cryIA(c)*	M73248	1–3531	83
cry1Ac6	*cryIA(c)*	U43606	1–>1821	63
cry1Ac7	*cryIA(c)*	U87793	976–4509	38
cry1Ac8	*cryIA(c)*	U87397	153–3686	71
cry1Ac9	*cryIA(c)*	U89872	388–3921	33
cry1Ac10		AJ002514	388–3921	107
cry1Ad1	*cryIA(c)*	M73250	1–3537	79
cry1Ae1	*cryIA(e)*	M65252	81–3623	60
cry1Af1	*icp*	U82003	172–>2905	49
cry1Ba1	*cryIB*	X06711	1–3684	10
cry1Ba2		X95704	186–3869	105
cry1Bb1	ET5	L32020	67–3753	25
cry1Bc1	*cryIB(c)*	Z46442	141–3839	6
cry1Bd1	*cryE1*	U70726		12
cry1Ca1	*cryIC*	X07518	47–3613	45
cry1Ca2	*cryIC*	X13620	241–>2711	88
cry1Ca3	*cryIC*	M73251	1–3570	79
cry1Ca4	*cryIC*	A27642	234–3800	114
cry1Ca5	*cryIC*	X96682	1–>2268	106
cry1Ca6	*cryIC*	X96683	1–>2268	106
cry1Ca7	*cryIC*	X96684	1–>2268	106
cry1Cb1	*cryIC(b)*	M97880	296–3823	48
cry1Da1	*cryID*	X54160	264–3758	42
cry1Db1	*prtB*	Z22511	241–3720	56
cry1Ea1	*cryIE*	X53985	130–3642	115
cry1Ea2	*cryIE*	X56144	1–3513	7
cry1Ea3	*cryIE*	M73252	1–3513	82
cry1Ea4		U94323	388–3900	47
cry1Eb1	*cryIE(b)*	M73253	1–3522	81
cry1Fa1	*cryIF*	M63897	478–3999	14
cry1Fa2	*cryIF*	M73254	1–3525	80
cry1Fb1	*prtD*	Z22512	483–4004	56
cry1Ga1	*prtA*	Z22510	67–3564	56
cry1Ga2	*cryIM*	Y09326	692–4210	96
cry1Gb1	*cryH2*	U70725		12
cry1Ha1	*prtC*	Z22513	530–4045	56
cry1Hb1		U35780	728–4195	53
cry1Ia1	*cryV*	X62821	355–2511	108
cry1Ia2	*cryV*	M98544	1–2157	34
cry1Ia3	*cryV*	L36338	279–2435	100
cry1Ia4	*cryV*	L49391	61–2217	54
cry1Ia5	*cryV159*	Y08920	524–2680	94
cry1Ib1	*cryV465*	U07642	237–2393	100
cry1Ja1	ET4	L32019	99–3519	25
cry1Jb1	ET1	U31527	177–3686	116
cry1Ka1		U28801	451–4098	52
cry2Aa1	*cryIIA*	M31738	156–2054	20
cry2Aa2	*cryIIA*	M23723	1840–3738	123
cry2Aa3		D86064	2007–3911	89
cry2Ab1	*cryIIB*	M23724	1–1899	123
cry2Ab2	*cryIIB*	X55416	874–2775	17
cry2Ac1	*cryIIC*	X57252	2125–3990	124
cry3Aa1	*cryIIIA*	M22472	25–1956	39
cry3Aa2	*cryIIIA*	J02978	241–2172	93
cry3Aa3	*cryIIIA*	Y00420	566–2497	41
cry3Aa4	*cryIIIA*	M30503	201–2132	65
cry3Aa5	*cryIIIA*	M37207	569–2500	22
cry3Aa6	*cryIIIA*	U10985	569–2500	1
cry3Ba1	*cryIIIB2*	X17123	25–>1977	101
cry3Ba2	*cryIIIB*	A07234	342–2297	85
cry3Bb1	*cryIIIBb*	M89794	202–2157	24
cry3Bb2	*cryIIIC(b)*	U31633	144–2099	23
cry3Ca1	*cryIIID*	X59797	232–2178	59
cry4Aa1	*cryIVA*	Y00423	1–3540	121
cry4Aa2	*cryIVA*	D00248	393–3935	95
cry4Ba1	*cryIVB*	X07423	157–3564	16
cry4Ba2	*cryIVB*	X07082	151–3558	112
cry4Ba3	*cryIVB*	M20242	526–3930	125
cry4Ba4	*cryIVB*	D00247	461–3865	95
cry5Aa1	*cryVA(a)*	L07025	1–>4155	102
cry5Ab1	*cryVA(b)*	L07026	1–>3867	67
cry5Ac1		I34543	1–>3660	76
cry5Ba1	PS86Q3	U19725	1–>3735	76
cry6Aa1	*cryVIA*	L07022	1–>1425	68
cry6Ba1	*cryVIB*	L07024	1–>1185	67
cry7Aa1	*cryIIIC*	M64478	184–3597	58
cry7Ab1	*cryIIIC(b)*	U04367	1–>3414	75
cry7Ab2	*cryIIIC(c)*	U04368	1–>3414	75
cry8Aa1	*cryIIIE*	U04364	1–>3471	29
cry8Ba1	*cryIIIG*	U04365	1–>3507	66
cry8Ca1	*cryIIIF*	U04366	1–3447	70
cry9Aa1	*cryIG*	X58120	5807–9274	104
cry9Aa2	*cryIG*	X58534	385–>3837	32
cry9Ba1	*cryX*	X75019	26–3488	97
cry9Ca1	*cryIH*	Z37527	2096–5569	57
cry9Da1	N141	D85560	47–3553	4
cry9Da2		AF042733	<1–>1937	122
cry10Aa1	*cryIVC*	M12662	941–2965	111
cry11Aa1	*cryIVD*	M31737	41–1969	21
cry11Aa2	*cryIVD*	M22860	<1–235	2
cry11Ba1	Jeg80	X86902	64–2238	19
cry11Bb1	94 kDa	AF017416		72
cry12Aa1	*cryVB*	L07027	1–>3771	67
cry13Aa1	*cryVC*	L07023	1–2409	90
cry14Aa1	*cryVD*	U13955	1–3558	77
cry15Aa1	34kDa	M76442	1036–2055	11
cry16Aa1	*cbm71*	X94146	158–1996	5
cry17Aa1	*cbm72*	X99478	12–1865	5
cry18Aa1	*cryBP1*	X99049	743–2860	126
cry19Aa1	Jeg65	Y07603	719–2662	86
cry19Ba1		D88381		87
cry20Aa1	86kDa	U82518	60–2318	61
cry21Aa1		I32932	1–3501	74
cry22Aa1		I34547	1–2169	76
cyt1Aa1	*cytA*	X03182	140–886	118
cyt1Aa2	*cytA*	X04338	509–1255	120
cyt1Aa3	*cytA*	Y00135	36–782	26
cyt1Aa4	*cytA*	M35968	67–813	30
cyt1Ab1	*cytM*	X98793	28–777	109
cyt1Ba1		U37196	1–795	78
cyt2Aa1	*cytB*	Z14147	270–1046	51
cyt2Ba1	*"cytB"*	U52043	287–655	35
cyt2Bb1		U82519	416–1204	15

[a] The symbols < and > indicate that the coding region extends up- or downstream, respectively, from the known sequence data.
[b] Only the polypeptide sequence has been reported.

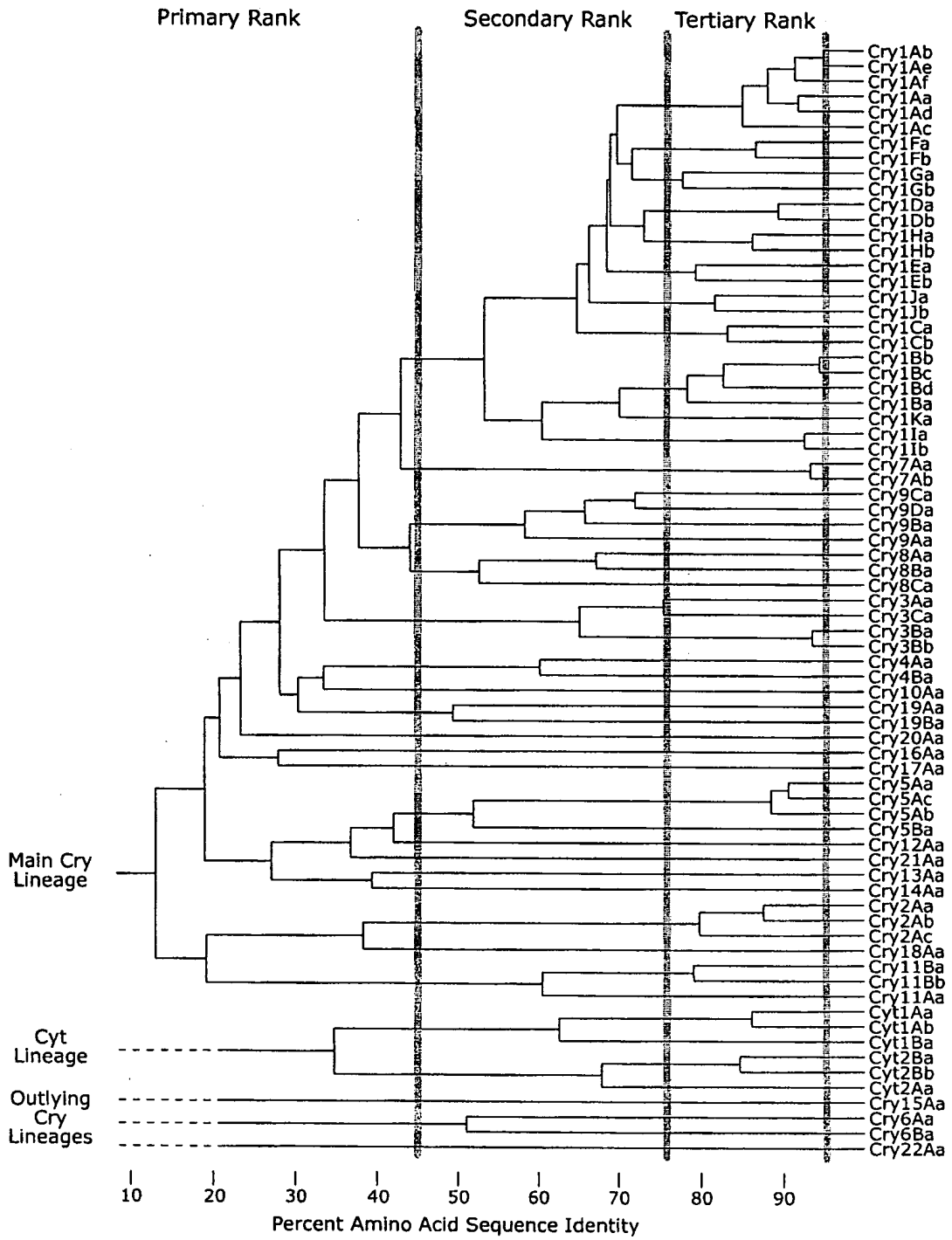


FIG. 1. Phylogram demonstrating amino acid sequence identity among Cry and Cyt proteins. This phylogenetic tree is modified from a TREEVIEW visualization of NEIGHBOR treatment of a CLUSTAL W multiple alignment and distance matrix of the full-length toxin sequences, as described in the text. The gray vertical bars demarcate the four levels of nomenclature ranks. Based on the low percentage of identical residues and the absence of any conserved sequence blocks in multiple-sequence alignments, the lower four lineages are not treated as part of the main toxin family, and their nodes have been replaced with dashed horizontal lines in this figure.

had significant branch points deeper in the tree than the chosen primary rank in the nomenclature. This sort of analysis was rejected as unsuitable for the purposes of Cry nomenclature due to the generally ragged branch lengths it produced and the requirement for the careful choice of an outgroup.

An alternative method of clustering protein sequences, capable of handling sequences that are quite diverse, is parsimony analysis. A consensus tree generated from 100 bootstraps of such an analysis displaces the two incomplete Cry1 sequences (Cry1Bd and Cry1Af) and the two Cry1 sequences lacking the C-terminal protoxin segments (Cry1Ia and Cry1Ib) into a region of the tree populated with such shortened se-

quences (not shown). With the further exceptions of Cry12A being interjected into the Cry5 cluster and a number of sequences besides Cry6B clustering higher in the tree than Cry6A, the proposed nomenclature successfully reflects the grouping of sequences provided by this method of analysis as well.

As noted above, the usual distance metrics for phylogenetic analysis account for multiple substitutions per site; most commonly, the Dayhoff PAM metric is used. When this distance metric was applied to the alignment used to make Fig. 1, a large number of the sequence pairs were found to have infinite distance. Therefore, the main Cry lineage and the Cyt lineage were separately aligned, the distances were calculated, and the distance matrices were clustered by using the FITCH program (of the PHYLIP software package). This method of analysis revealed several strongly associated groups of sequences (>90% of trees) in the main Cry lineage that extend deeper into the tree than the primary rank assigned in the proposed nomenclature: Cry1; Cry3; Cry4; Cry7; the Cry5, Cry12-Cry13-Cry14-Cry21 group; the Cry8-Cry9 group; the Cry10-Cry19 group; the Cry16-Cry17 group; and the Cry2-Cry11-Cry18 group. Many of these groups, however, were separated by branch points that were either nonmajority or were found <60% of the time; thus, the arrangement of these groups would be likely to change with additional sequence additions. At the secondary rank, the only anomaly with respect to the proposed nomenclature was the interjection of the Cry1Ia and Cry1Ib sequences into the Cry1B group. This effect may be due to an artificially reduced distance between the Cry1I sequences and the incomplete Cry1Bd sequence caused by the particular distance metric used. The Cyt lineage sequences were separated into the expected two primary rank groups that separate into the expected secondary rank groupings. This more standard phylogenetic approach also suffers from an accentuated visual disorientation of uneven branch lengths and shortening of the more closely related branches, especially at the tertiary rank (lowercase letter), where a great deal of comparative work has been done among the Cry1 toxins.

In summary, the proposed nomenclature uses readily available software that can be easily interpreted by investigators in the field and meets their needs as well as, or better than, alternative methods of analysis and presentation. When the holotype toxins were analyzed by alternative phylogenetic methods, the hierarchy implied by the nomenclature was essentially consistent with the resulting phylogenetic clustering, and the few exceptions were largely explainable by known properties of the sequences in question.

ACKNOWLEDGMENTS

The BGSC is supported by National Science Foundation grant DBI-9319712 and by industrial sponsorships.

REFERENCES

1. **Adams, L. F., S. Mathewes, P. O'Hara, A. Petersen, and H. Gürtler.** 1994. Elucidation of the mechanism of CryIIIA overproduction in a mutagenized strain of *Bacillus thuringiensis* var. *tenebrionis*. Mol. Microbiol. **14**:381–389.
2. **Adams,L. F., J. E. Visick, and H. R. Whiteley.** 1989. A 20-kilodalton protein is required for efficient production of the *Bacillus thuringiensis* subsp. *israelensis* 27-kilodalton crystal protein in *Escherichia coli*. J. Bacteriol. **171**: 521–530.
3. **Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker, and D. V. Thompson.** 1985. Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. Gene **36**:289–300.
4. **Asano, S. I., Y. Nukumizu, H. Bando, T. Iizuka, and T. Yamamoto.** 1997. Cloning of novel enterotoxin genes from *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. **63**:1054–1057.
5. **Barloy, F., A. Delécluse, L. Nicolas, and M.-M. Lecadet.** 1996. Cloning and expression of the first anaerobic toxin gene from *Clostridium bifermentans* subsp. *malaysia*, encoding a new mosquitocidal protein with homologies to *Bacillus thuringiensis* delta-endotoxins. J. Bacteriol. **178**:3099–3105.
6. **Bishop, A. H.** 1994. Unpublished observation.
7. **Bossé, M., L. Masson, and R. Brousseau.** 1990. Nucleotide sequence of a novel crystal protein gene isolated from *Bacillus thuringiensis* subspecies *kenyae*. Nucleic Acids Res. **18**:7443.
8. **Bradley, D., M. A. Harkey, M.-K. Kim, D. Biever, and L. S. Bauer.** 1995. The insecticidal CryIB protein of *Bacillus thuringiensis* ssp. *thuringiensis* has dual specificity to coleopteran and lepidopteran larvae. J. Invertebr. Pathol. **65**:162–173.
9. **Bravo, A.** 1997. Phylogenetic relationships of *Bacillus thuringiensis* δ-endotoxin family proteins and their functional domains. J. Bacteriol. **179**:2793–2801.
10. **Brizzard, B. L., and H. R. Whiteley.** 1988. Nucleotide sequence of an additional crystal protein gene cloned from *Bacillus thuringiensis* subsp. *thuringiensis*. Nucleic Acids Res. **16**:2723–2724.
11. **Brown, K. L., and H. R. Whiteley.** 1992. Molecular characterization of two novel crystal protein genes from *Bacillus thuringiensis* subsp. *thompsoni*. J. Bacteriol. **174**:549–557.
12. **Chak, K. F.** 1996. Unpublished observation.
13. **Chak, K. F., and J. C. Chen.** 1993. Complete nucleotide sequence and identification of a putative promoter region for the expression in *Escherichia coli* of the *cryIA(b)* gene from *Bacillus thuringiensis* var. *aizawai* HD133. Proc. Natl. Sci. Counc. Repub. China **17**:7–14.
14. **Chambers, J. A., A. Jelen, M. P. Gilbert, C. S. Jany, T. B. Johnson, and C. Gawron-Burke.** 1991. Isolation and characterization of a novel insecticidal crystal protein gene from *Bacillus thuringiensis* subsp. *aizawai*. J. Bacteriol. **173**:3966–3976.
15. **Cheong, H., and S. S. Gill.** 1997. Cloning and characterization of a cytolytic and mosquitocidal δ-endotoxin from *Bacillus thuringiensis* subsp. *jegathesan*. Appl. Environ. Microbiol. **63**:3254–3260.
16. **Chungjatupornchai, W., H. Höfte, J. Seurinck, C. Angsuthanasombat, and M. Vaeck.** 1988. Common features of *Bacillus thuringiensis* toxins specific for *Diptera* and *Lepidoptera*. Eur. J. Biochem. **173**:9–16.
17. **Dankocsik, C., W. P. Donovan, and C. S. Jany.** 1990. Activation of a cryptic crystal protein gene of *Bacillus thuringiensis* subspecies *kurstaki* by gene fusion and determination of the crystal protein insecticidal specificity. Mol. Microbiol. **4**:2087–2094.
18. **Dardenne, F., J. Seurinck, B. Lambert, and M. Peferoen.** 1990. Nucleotide sequence and deduced amino acid sequence of a *cryIA(c)* gene variant from *Bacillus thuringiensis*. Nucleic Acids Res. **18**:5546.
19. **Delécluse, A., M.-L. Rosso, and A. Ragni.** 1995. Cloning and expression of a novel toxin gene from *Bacillus thuringiensis* subsp. *jegathesan* encoding a highly mosquitocidal protein. Appl. Environ. Microbiol. **61**:4230–4235.
20. **Donovan, W. P., C. C. Dankocsik, M. P. Gilbert, W. C. Gawron-Burke, R. R. Groat, and B. C. Carlton.** 1988. Amino acid sequence and entomocidal activity of the P2 crystal protein. An insect toxin from *Bacillus thuringiensis* var. *kurstaki*. J. Biol. Chem. **263**:561–567. (Author's correction, **263**:4740.)
21. **Donovan, W. P., C. Dankocsik, and M. P. Gilbert.** 1988. Molecular characterization of a gene encoding a 72-kilodalton mosquito-toxic crystal protein from *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. **170**:4732–4738.
22. **Donovan, W. P., J. M. González, Jr., M. P. Gilbert, and C. Dankocsik.** 1988. Isolation and characterization of EG2158, a new strain of *Bacillus thuringiensis* toxic to coleopteran larvae, and nucleotide sequence of the toxin gene. Mol. Gen. Genet. **214**:365–372.
23. **Donovan, W. P., M. J. Rupar, and A. C. Slaney.** January 1995. U.S. patent 5,378,625.
24. **Donovan, W. P., M. J. Rupar, A. C. Slaney, T. Malvar, M. C. Gawron-Burke, and T. B. Johnson.** 1992. Characterization of two genes encoding *Bacillus thuringiensis* insecticidal crystal proteins toxic to *Coleoptera* species. Appl. Environ. Microbiol. **58**:3921–3927.
25. **Donovan, W. P., Y. Tan, C. S. Jany, and J. M. González, Jr.** June 1994. U.S. patent 5,322,687.
26. **Earp, D. J., and D. J. Ellar.** 1987. *Bacillus thuringiensis* var. *morrisoni* strain PG14: nucleotide sequence of a gene encoding a 27 kDa crystal protein. Nucleic Acids Res. **15**:3619.
27. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 2). Cladistics **5**:164–166.
28. **Fischhoff, D. A., K. S. Bowdisch, F. J. Perlak, P. G. Marrone, S. H. McCormick, J. G. Niedermeyer, D. A. Dean, K. Kusano-Kretzmer, E. J. Mayer, D. E. Rochester, S. G. Rogers, and R. T. Fraley.** 1987. Insect tolerant transgenic tomato plants. Bio/Technology **5**:807–813.
29. **Foncerrada, L., A. J. Sick, and J. M. Payne.** August 1992. European Patent Office no. EP 0498537.
30. **Galjart, N. J., N. Sivasubramanian, and B. A. Federici.** 1987. Plasmid location, cloning and sequence analysis of the gene encoding a 23-kilodalton cytolytic protein from *Bacillus thuringiensis* subsp. *morrisoni* (PG-14). Curr. Microbiol. **16**:171–177.
31. **Geiser, M., S. Schweitzer, and C. Grimm.** 1986. The hypervariable region in the genes coding for entomopathogenic crystal proteins of *Bacillus thurin-*

giensis: nucleotide sequence of the kurhd1 gene of subsp. *kurstaki* HD1. Gene **48**:109–118.
32. **Gleave, A. P., R. J. Hedges, and A. H. Broadwell.** 1992. Identification of an insecticidal crystal protein from *Bacillus thuringiensis* DSIR517 with significant sequence differences from previously described toxins. J. Gen. Microbiol. **138**:55–62.
33. **Gleave, A. P., R. J. Hedges, A. H. Broadwell, and P. J. Wigley.** 1992. Cloning and nucleotide sequence of an insecticidal crystal protein gene from *Bacillus thuringiensis* DSIR732 active against three species of leafroller Lepidoptera *Tortricidae*. N. Z. J. Crop Hortic. Sci. **20**:27–36.
34. **Gleave, A. P., R. Williams, and R. J. Hedges.** 1993. Screening by polymerase chain reaction of *Bacillus thuringiensis* serotypes for the presence of *cryV*-like insecticidal protein genes and characterization of a *cryV* gene cloned from *B. thuringiensis* subsp. *kurstaki*. Appl. Environ. Microbiol. **59**:1683–1687.
35. **Guerchicoff, A., R. U. Ugalde, and C. P. Rubinstein.** 1997. Identification and characterization of a previously undescribed *cyt* gene in *Bacillus thuringiensis* subsp. *israelensis*. Appl. Environ. Microbiol. **63**:2716–2721.
36. **Haider, M. Z., and D. J. Ellar.** 1988. Nucleotide sequence of a *Bacillus thuringiensis aizawai* ICI entomocidal crystal protein gene. Nucleic Acids Res. **16**:10927.
37. **Hefford, M. A., R. Brousseau, G. Préfontaine, Z. Hanna, J. A. Condie, and P. C. K. Lau.** 1987. Sequence of a lepidopteran toxin gene of *Bacillus thuringiensis* subsp. *kurstaki* NRD-12. J. Biotechnol. **6**:307–322.
38. **Herrera, G., S. J. Snyman, and J. A. Thomson.** 1994. Construction of a bioinsecticidal strain of *Pseudomonas flourescens* active against the sugarcane borer, *Eldana saccharina*. Appl. Environ. Microbiol. **60**:682–690.
39. **Herrnstadt, C., T. E. Gilroy, D. A. Sobieski, B. D. Bennett, and F. H. Gaertner.** 1987. Nucleotide sequence and deduced amino acid sequence of a coleopteran-active delta-endotoxin gene from *Bacillus thuringiensis* subsp. *san diego*. Gene **57**:37–46.
40. **Höfte, H., H. de Greve, J. Seurinck, S. Jansens, J. Mahillon, C. Ampe, J. Vandekerckhove, M. van Montagu, M. Zabeau, and M. Vaeck.** 1986. Structural and functional analysis of a cloned delta endotoxin of *Bacillus thuringiensis berliner* 1715. Eur. J. Biochem. **161**:273–280.
41. **Höfte, H., J. Seurinck, A. Van Houtven, and M. Vaeck.** 1987. Nucleotide sequence of a gene encoding an insecticidal protein of *Bacillus thuringiensis* var. *tenebrionis* toxic against Coleoptera. Nucleic Acids Res. **15**:7183.
42. **Höfte, H., P. Soetaert, S. Jansens, and M. Peferoen.** 1990. Nucleotide sequence and deduced amino acid sequence of a new Lepidoptera-specific crystal protein gene from *Bacillus thuringiensis*. Nucleic Acids Res. **18**:5545.
43. **Höfte, H., J. Van Rie, S. Jansens, A. Van Houtven, H. Vanderbruggen, and M. Vaeck.** 1988. Monoclonal antibody analysis and insecticidal spectrum of three types of lepidopteran-specific insecticidal crystal proteins of *Bacillus thuringiensis*. Appl. Environ. Microbiol. **54**:2010–2017.
44. **Höfte, H., and H. R. Whiteley.** 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. **53**:242–255.
45. **Honée, G., T. van der Salm, and B. Visser.** 1988. Nucleotide sequence of crystal protein gene isolated from *B. thuringiensis* subspecies *entomocidus* 60.5 coding for a toxin highly active against *Spodoptera* species. Nucleic Acids Res. **16**:6240.
46. **Hori, H., K. Suzuki, K. Ogiwara, M. Minani, M. Himejima, K. Sakanaka, Y. Kaji, S. Asano, R. Sato, M. Ohba, and H. Iwahana.** 1992. Presented at the XXVth Annual Meeting of the Society for Invertebrate Pathology, Heidelberg, Germany.
47. **Ibarra, J.** 1997. Unpublished observation.
48. **Kalman, S., K. L. Kiehne, J. L. Libs, and T. Yamamoto.** 1993. Cloning of a novel *cryIC*-type gene from a strain of *Bacillus thuringiensis* subsp. *galleriae*. Appl. Environ. Microbiol. **59**:1131–1137.
49. **Kang, S. K., H. S. Kim, and Y. M. Yu.** 1997. Unpublished observation.
50. **Kondo, S., N. Tamura, A. Kunitate, M. Hattori, A. Akashi, and I. Ohmori.** 1987. Cloning and nucleotide sequencing of two insecticidal δ-endotoxin genes from *Bacillus thuringiensis* var. *kurstaki* HD-1 DNA. Agric. Biol. Chem. **51**:455–463.
51. **Koni, P. A., and D. J. Ellar.** 1993. Cloning and characterization of a novel *Bacillus thuringiensis* cytolytic delta-endotoxin. J. Mol. Biol. **229**:319–327.
52. **Koo, B. T.** 1995. Unpublished observation.
53. **Koo, B. T., S. H. Park, S. K. Choi, B. S. Shin, J. I. Kim, and J. H. Yu.** 1995. Cloning of a novel crystal protein gene *cry1K* from *Bacillus thuringiensis* subsp *morrisoni*. FEMS Microbiol. Lett. **134**:159–164.
54. **Kostichka, K., G. W. Warren, M. Mullins, A. D. Mullins, J. A. Craig, M. G. Koziel, and J. J. Estruch.** 1996. Cloning of a *cryV*-type insecticidal protein gene from *Bacillus thuringiensis*: the *cryV*-encoded protein is expressed early in stationary phase. J. Bacteriol. **178**:2141–2144.
55. **Kronstad, J. W., and H. R. Whiteley.** 1986. Three classes of homologous *Bacillus thuringiensis* crystal-protein genes. Gene **43**:29–40.
56. **Lambert, B.** 1993. Unpublished observation.
57. **Lambert, B., L. Buysse, C. Decock, S. Jansens, C. Piens, B. Saey, J. Seurinck, K. Van Audenhove, J. Van Rie, A. Van Vliet, and M. Peferoen.** 1996. A *Bacillus thuringiensis* insecticidal protein with a high activity against members of the family Noctuidae. Appl. Environ. Microbiol. **62**:80–86.
58. **Lambert, B., H. Höfte, K. Annys, S. Jansens, P. Soetaert, and M. Peferoen.** 1992. Novel *Bacillus thuringiensis* insecticidal crystal protein with a silent activity against coleopteran larvae. Appl. Environ. Microbiol. **58**:2536–2542.
59. **Lambert, B., W. Theunis, R. Agouda, K. Van Audenhove, D. C., S. Jansens, J. Seurinck, and M. Peferoen.** 1992. Nucleotide sequence of gene *cryIIID* encoding a novel coleopteran-active crystal protein from strain BTI109P of *Bacillus thuringiensis* subsp. *kurstaki*. Gene **110**:131–132.
60. **Lee, C.-S., and A. I. Aronson.** 1991. Cloning and analysis of δ-endotoxin genes from *Bacillus thuringiensis* subsp. *alesti*. J. Bacteriol. **173**:6635–6638.
61. **Lee, H.-K., and S. S. Gill.** 1997. Molecular cloning and characterization of a novel mosquitocidal protein gene from *Bacillus thuringiensis* subsp. *fukuokaensis*. Appl. Environ. Microbiol. **63**:4664–4670.
62. **Masson, L., P. Marcotte, G. Préfontaine, and R. Brousseau.** 1989. Nucleotide sequence of a gene cloned from *Bacillus thuringiensis* subspecies *entomocidus* coding for an insecticidal protein toxic for *Bombyx mori*. Nucleic Acids Res. **17**:446.
63. **Masson, L., A. Mazza, L. Gringorten, D. Baines, V. Aneliunas, and R. Brousseau.** 1994. Specificity domain localization of *Bacillus thuringiensis* insecticidal toxins is highly dependent on the bioassay system. Mol. Microbiol. **14**:851–860.
64. **McLinden, J. H., J. R. Sabourin, B. D. Clark, D. R. Gensler, W. E. Workman, and D. H. Dean.** 1985. Cloning and expression of an insecticidal k-73 type crystal protein gene from *Bacillus thuringiensis* var. *kurstaki* into *Escherichia coli*. Appl. Environ. Microbiol. **50**:623–628.
65. **McPherson, S. A., F. J. Perlak, R. L. Fuchs, P. G. Marrone, P. B. Lavrik, and D. A. Fischhoff.** 1988. Characterization of the coleopteran-specific protein gene of *Bacillus thuringiensis* var. *tenebrionis*. Bio/Technology **6**:61–66.
66. **Michaels, T. E., K. E. Narva, and L. Foncerrada.** August 1993. World Intellectual Property Organization patent WO 93/15206.
67. **Narva, K. E., J. M. Payne, G. E. Schwab, L. A. Hickle, T. Galasan, and A. J. Sick.** December 1991. European Patent Office no. EP 0462721.
68. **Narva, K. E., G. E. Schwab, T. Galasan, and J. M. Payne.** August 1993. U.S. patent 5,236,843.
68a. **Nelson, D. R., L. Koymans, T. Kamataki, J. J. Stegeman, R. Feyereisen, D. J. Waxman, M. R. Waterman, O. Gotoh, M. J. Coon, R. W. Estabrook, I. C. Gunsalus, and D. W. Nebert.** 1996. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics **6**:1–42.
69. **Oeda, K., K. Oshie, M. Shimizu, K. Nakamura, H. Yamamoto, I. Nakayama, and H. Ohkawa.** 1987. Nucleotide sequence of the insecticidal protein gene of *Bacillus thuringiensis* strain *aizawai* IPL7 and its high-level expression in *Escherichia coli*. Gene **53**:113–119.
70. **Ogiwara, K., H. Hori, M. Minami, K. Takeuchi, R. Sato, M. Ohba, and H. Iwahana.** 1995. Nucleotide sequence of the gene encoding novel delta-endotoxin from *Bacillus thuringiensis* serovar *japonensis* strain Buibui specific to scarabaeid beetles. Curr. Microbiol. **30**:227–235.
71. **Omolo, E. O., J. M. D., O. E. O., and J. A. Thomson.** 1997. Cloning and expression of a *Bacillus thuringiensis* (L1-2) gene encoding a crystal protein active against *Glossina morsitans morsitans* and *Chilo partellus*. Curr. Microbiol. **34**:118–121.
72. **Orduz, S.** Unpublished observation.
73. **Page, R. D. M.** 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. CABIOS **12**:357–358.
74. **Payne, J., K. E. Narva, and J. Fu.** December 1996. U.S. patent 5,589,382.
75. **Payne, J. M., and J. M. Fu.** February 1994. U.S. patent 5,286,486.
76. **Payne, J. M., M. K. Kennedy, J. B. Randall, H. Meier, H. J. Uick, L. Foncerrada, H. E. Schnepf, G. E. Schwab, and J. Fu.** January 1997. U.S. patent 5,596,071.
77. **Payne, J. M., and K. E. Narva.** July 1994. World Intellectual Property Organization patent WO 94/16079.
78. **Payne, J. M., K. E. Narva, K. A. Uyeda, C. J. Stalder, and T. E. Michaels.** July 1995. U.S. patent 5,436,002.
79. **Payne, J. M., and A. J. Sick.** September 1993. U.S. patent 5,246,852.
80. **Payne, J. M., and A. J. Sick.** February 1993. U.S. patent 5,188,960.
81. **Payne, J. M., and A. J. Sick.** April 1993. U.S. patent 5,206,166.
82. **Payne, J. M., and A. J. Sick.** August 1991. U.S. patent 5,039,523.
83. **Payne, J. M., A. J. Sick, and M. Thompson.** August 1992. U.S. patent 5,135,867.
84. **Payne, J. M., G. G. Soares, H. W. Talbot, and T. C. Olson.** October 1991. U.S. patent 4,990,332.
85. **Peferoen, M., B. Lambert, and H. Joos.** August 1990. European patent Office no. EP 0382990-A1.
86. **Rosso, M. L., and A. Delecluse.** 1997. Contribution of the 65-kilodalton protein encoded by the cloned gene *cry19A* to the mosquitocidal activity of *Bacillus thuringiensis* subsp. *jegathesan*. Appl. Environ. Microbiol. **63**:4449–4455.
87. **Saitoh, H.** 1996. Unpublished observation.
88. **Sanchis, V., D. Lereclus, G. Menou, J. Chaufaux, S. Guo, and M.-M. Lecadet.** 1989. Nucleotide sequence and analysis of the N-terminal coding region of the *Spodoptera*-active δ-endotoxin gene of *Bacillus thuringiensis aizawai* 7.29. Mol. Microbiol. **3**:229–238.

89. **Sasaki, J., S. Asano, N. Hashimoto, B.-W. Lay, S. Hastowo, H. Bando, and T. Iizuka.** 1997. Characterization of a *cry2A* gene cloned from an isolate of *Bacillus thuringiensis* serovar *sotto*. Curr. Microbiol. **35**:1–8.
90. **Schnepf, H. E., G. E. Schwab, J. M. Payne, K. E. Narva, and L. Foncerrada.** November 1992. World Intellectual Property Organization patent WO 92/19739.
91. **Schnepf, H. E., and H. R. Whiteley.** 1981. Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:2893–2897.
92. **Schnepf, H. E., H. C. Wong, and H. R. Whiteley.** 1985. The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. J. Biol. Chem. **260**:6264–6272.
93. **Sekar, V., D. V. Thompson, M. J. Maroney, R. G. Bookland, and M. J. Adang.** 1987. Molecular cloning and characterization of the insecticidal crystal protein gene of *Bacillus thuringiensis* var. *tenebrionis*. Proc. Natl. Acad. Sci. USA **84**:7036–7040.
94. **Selvapandiyan, A.** 1996. Unpublished observation.
95. **Sen, K., G. Honda, N. Koyama, M. Nishida, A. Neki, H. Sakai, M. Himeno, and T. Komano.** 1988. Cloning and nucleotide sequences of the two 130 kDa insecticidal protein genes of *Bacillus thuringiensis* var. *israelensis*. Agric. Biol. Chem. **52**:873–878.
96. **Shevelev, A. B., Y. N. Kogan, A. M. Busheva, E. J. Voronina, D. V. Tebrikov, S. I. Novikova, G. G. Chestukhina, V. Kubshinov, E. Pehu, and V. M. Stepanov.** 1997. A novel delta-endotoxin gene *cryIM* from *Bacillus thuringiensis* subsp. *wuhanensis*. FEBS Lett. **404**:148–152.
97. **Shevelev, A. B., M. A. Svarinsky, A. I. Karasin, Y. N. Kogan, G. G. Chestukhina, and V. M. Stepanov.** 1993. Primary structure of *cryX*, the novel delta-endotoxin-related gene from *Bacillus thuringiensis* spp. *galleriae*. FEBS Lett. **336**:79–82.
98. **Shibano, Y., A. Yamagata, N. Nakamura, T. Iizuka, H. Sugisaki, and M. Takanami.** 1985. Nucleotide sequence coding for the insecticidal fragment of the *Bacillus thuringiensis* crystal protein. Gene **34**:243–251.
99. **Shimizu, M., K. Oshie, K. Nakamura, Y. Takada, K. Oeda, and H. Ohkawa.** 1988. Cloning and expression in *Escherichia coli* of the 135-kDa insecticidal protein gene from *Bacillus thuringiensis* subsp. *aizawai* IPL7. Agric. Biol. Chem. **52**:1565–1573.
100. **Shin, B.-S., S.-H. Park, S.-K. Choi, B.-T. Koo, S.-T. Lee, and J.-I. Kim.** 1995. Distribution of *cryV*-type insecticidal protein genes in *Bacillus thuringiensis* and cloning of *cryV*-type genes from *Bacillus thuringiensis* subsp. *kurstaki* and *Bacillus thuringiensis* subsp. *entomocidus*. Appl. Environ. Microbiol. **61**:2402–2407.
101. **Sick, A., F. Gaertner, and A. Wong.** 1990. Nucleotide sequence of a coleopteran-active toxin gene from a new isolate of *Bacillus thuringiensis* subsp. *tolworthi*. Nucleic Acids Res. **18**:1305.
102. **Sick, A. J., G. E. Schwab, and J. M. Payne.** January 1994. U.S. patent 05281530.
103. **Smith, G. P., and D. J. Ellar.** 1994. Mutagenesis of two surface-exposed loops of the *Bacillus thuringiensis* CryIC δ-endotoxin affects insecticidal specificity. Biochem. J. **302**:611–616.
104. **Smulevitch, S. V., A. L. Osterman, A. B. Shevelev, S. V. Kaluger, A. I. Karasin, R. M. Kadyrov, O. P. Zagnitko, G. G. Chestukhina, and V. M. Stepanov.** 1991. Nucleotide sequence of a novel delta-endotoxin gene *cryIG* of *Bacillus thuringiensis* ssp. *galleriae*. FEBS Lett. **293**:25–28.
105. **Soetaert, P.** 1996. Unpublished observation.
106. **Strizhov, N.** 1996. Unpublished observation.
107. **Sun, M.** 1997. Unpublished observation.
108. **Tailor, R., J. Tippett, G. Gibb, S. Pells, D. Pike, L. Jordon, and S. Ely.** 1992. Identification and characterization of a novel *Bacillus thuringiensis* delta-endotoxin entomocidal to coleopteran and lepidopteran larvae. Mol. Microbiol. **6**:1211–1217.
109. **Thiery, I., A. Delécluse, M. C. Tamayo, and S. Orduz.** 1997. Identification of a gene for Cyt1A-like hemolysin from *Bacillus thuringiensis* subsp. *medellin* and expression in a crystal-negative *B. thuringiensis* strain. Appl. Environ. Microbiol. **63**:468–473.
110. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.
111. **Thorne, L., F. Garduno, T. Thompson, D. Decker, M. Zounes, M. Wild, A. M. Walfield, and T. J. Pollock.** 1986. Structural similarity between the Lepidoptera- and Diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. "*kurstaki*" and "*israelensis*." J. Bacteriol. **166**:801–811.
112. **Tungpradubkul, S., C. Settasatien, and S. Panyim.** 1988. The complete nucleotide sequence of a 130 kDa mosquito-larvicidal delta-endotoxin gene of *Bacillus thuringiensis* var. *israelensis*. Nucleic Acids Res. **16**:1637–1638.
113. **Udayasuriyan, V., A. Nakamura, H. Mori, H. Masaki, and T. Uozumi.** 1994. Cloning of a new *cryIA(a)* gene from *Bacillus thuringiensis* strain FU-2-7 and analysis of chimeric *cryIA(a)* proteins for toxicity. Biosci. Biotechnol. Biochem. **58**:830–835.
114. **Van Mellaert, H., J. Botterman, J. Van Rie, and H. Joos.** January 1991. European Patent EP Office no. 0408403.
115. **Visser, B., E. Munsterman, A. Stoker, and W. G. Dirkse.** 1990. A novel *Bacillus thuringiensis* gene encoding a *Spodoptera exigua*-specific crystal protein. J. Bacteriol. **172**:6783–6788.
116. **Von Tersch, M. A., and J. M. Gonzalez.** October 1994. U.S. patent 5,356,623.
117. **Von Tersch, M. A., H. L. Robbins, C. S. Jany, and T. B. Johnson.** 1991. Insecticidal toxins from *Bacillus thuringiensis* subsp. *kenyae*: gene cloning and characterization and comparison with *B. thuringiensis* subsp. *kurstaki* CryIA(c) toxins. Appl. Environ. Microbiol. **57**:349–358.
118. **Waalwijk, C., A. M. Dullemans, M. E. S. vanWorkum, and B. Visser.** 1985. Molecular cloning and the nucleotide sequence of the Mr28,000 crystal protein gene of *Bacillus thuringiensis* subsp. *israelensis*. Nucleic Acids Res. **13**:8207–8217.
119. **Wabiko, H., K. C. Raymond, and L. A. Bulla, Jr.** 1986. *Bacillus thuringiensis* entomocidal protoxin gene sequence and gene product analysis. DNA **5**:305–314.
120. **Ward, E. S., and D. J. Ellar.** 1986. *Bacillus thuringiensis* var. *israelensis* delta-endotoxin: nucleotide sequence and characterization of the transcripts in *Bacillus thuringiensis* and *Escherichia coli*. J. Mol. Biol. **191**:1–11.
121. **Ward, E. S., and D. J. Ellar.** 1987. Nucleotide sequence of a *Bacillus thuringiensis* var. *israelensis* gene encoding a 130 kDa delta-endotoxin. Nucleic Acids Res. **15**:7195.
122. **Wasano, N., and M. Ohba.** 1998. Unpublished observation.
122a. **White, J. A., L. J. Maltais, and D. W. Nebert.** 1998. An increasingly urgent need for standardized gene nomenclature. See: http://genetics.nature.com/web_specials/nomen/nomen_article.html.
123. **Widner, W. R., and H. R. Whiteley.** 1989. Two highly related insecticidal crystal proteins of *Bacillus thuringiensis* subsp. *kurstaki* possess different host range specificities. J. Bacteriol. **171**:965–974.
124. **Wu, D., X. L. Cao, Y. Y. Bai, and A. I. Aronson.** 1991. Sequence of an operon containing a novel δ-endotoxin gene from *Bacillus thuringiensis*. FEMS Microbiol. Lett. **81**:31–36.
125. **Yamamoto, T., I. A. Watkinson, L. Kim, M. V. Sage, R. Stratton, N. Akande, Y. Li, D.-P. Ma, and B. A. Roe.** 1988. Nucleotide sequence of the gene coding for a 130-kDa mosquitocidal protein of *Bacillus thuringiensis israelensis*. Gene **66**:107–120.
126. **Zhang, J., T. C. Hodgman, L. Krieger, W. Schnetter, and H. U. Schairer.** 1997. Cloning and analysis of the first *cry* gene from *Bacillus popilliae*. J. Bacteriol. **179**:4336–4341.

Editor's note: Articles published in this journal represent the opinions of the authors and do not necessarily represent the opinions of ASM.

42. Porter, N. A. & Weekes, T. C. *Mon. Not. R. astr. Soc.* **183**, 205–210 (1978).
43. Akerlof, C. W. *et al.* in *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 135–138 (Adelaide University, 1990).
44. Porter, N. A. & Weekes, T. C. *Astrophys. J.* **212**, 224–226 (1977).
45. Porter, N. A. & Weekes, T. C. *Mon. Not. R. Astr. Soc.* **183**, 205–210 (1978).
46. Porter, N. A. & Weekes, T. C. *Nature* **277**, 199 (1979).
47. Nolan, K., Porter, N. A., Fegan, D. J., Chantell, M. & Weekes, T. C. *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 150 (Adelaide University, 1990).
48. Fegan, D. J., McBreen, B., O'Brien, D. & O'Sullivan, C. *Nature* **271**, 731–732 (1978).
49. Bhat, P. N. *et al. Nature* **284**, 433–434 (1984).
50. Porter, N. A. & Weekes, T. C. *Nature* **267**, 500–501 (1977).
51. O'Mongain, E. *Nature* **242**, 136–137 (1973).
52. Huguenin, G. R. & Moore, E. L. *Astrophys. J.* **187**, L57–L58 (1974).
53. Phinney, S. & Taylor, J. H. *Nature* **277**, 117–118 (1979).
54. Rees, M. J. *Nature* **266**, 333–334 (1977).
55. Jelley, J. V., Baird, G. A. & O'Mongain, E. *Nature* **267**, 499–500 (1977).

ACKNOWLEDGEMENTS. We thank C. J. Goebel for giving his time whenever we needed expert help. We thank N. Porter for discussions. This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by the US Department of Energy, in part by the Xunta de Galicia (Spain) and in part by the Smithsonian Scholarly Studies Research Fund. J.H.M. is a NAS/NRC research associate.

ARTICLES

Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution

Jade Li*, Joe Carroll† & David J. Ellar†

* Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Biochemistry Department, Cambridge University, Tennis Court Road, Cambridge CB2 1QW, UK

The structure of the δ-endotoxin from *Bacillus thuringiensis* subsp. *tenebrionis* that is specifically toxic to Coleoptera insects (beetle toxin) has been determined at 2.5 Å resolution. It comprises three domains which are, from the N- to C-termini, a seven-helix bundle, a three-sheet domain, and a β sandwich. The core of the molecule encompassing all the domain interfaces is built from conserved sequence segments of the active δ-endotoxins. Therefore the structure represents the general fold of this family of insecticidal proteins. The bundle of long, hydrophobic and amphipathic helices is equipped for pore formation in the insect membrane, and regions of the three-sheet domain are probably responsible for receptor binding.

THE δ-endotoxins are a family of insecticidal proteins produced by *Bacillus thuringiensis* (B.t.) during sporulation, having relative molecular masses (M_r) 60,000–70,000 (60K–70K) in the active form and specific toxicities against insects in the orders of Lepidoptera, Diptera and Coleoptera[1,2]. These toxins have been formulated into commercial insecticides for three decades[3], and now insect-resistant plants are engineered by transformation with Lepidoptera-specific toxin genes[4–6]. In the bacterium δ-endotoxins are synthesized as protoxins of M_rs 70K–135K and crystallize as a parasporal inclusion ~1 μ in size, in which form they are ingested by the susceptible insect. The microcrystal dissolves in the alkaline pH of the midgut and the protoxin is cleaved by gut proteases to release the active toxin. δ-Endotoxins activated *in vitro* bind specifically and with high affinity ($k_D \approx$ 0.1–20 nM) to protein receptors on brush-border membrane vesicles derived from the gut epithelium of target insects[7–9] and create leakage channels of 10–20 Å diameter in the cell membrane[10]. *In vivo* such membrane lesions lead to swelling and lysis of the gut epithelium[11] and death of the insect ensues through starvation and septicaemia. Active δ-endotoxins of different specificities show five strongly conserved regions in their amino-acid sequences[1,12]. Exchanging sequence segments in the divergent regions between toxins of different specificities can produce active hybrids showing altered target specificity[13–15]. We have determined the atomic structure of a Coleoptera-specific δ-endotoxin (CryIIIA, beetle toxin) from *B.t.* subsp. *tenebrionis*[16–18] to elucidate the structural basis for target specificity and membrane perforation by this family of proteins.

Structure determination

Parasporal crystals of the beetle toxin contain the full-length 644-residue protoxin[17] as the minor component, and a product of bacterial processing with 57 residues removed from the N-terminus as the major component[19]. The latter (M_r 67K) is similar in sequence to the active form of other δ-endotoxins. After solubilization, papain cleavage converts the mixture to the 67K toxin (see legend to Table 1). This was recrystallized in the original crystal form of the parasporal crystals, space group $C222_1$ and cell dimensions 117.1 by 134.2 by 104.5 Å, containing one molecule per asymmetric unit and 55% solvent by volume[18].

Initial evaluation of derivatives was carried out at 4.5 Å resolution with data collected on the FAST TV diffractometer[20] using CuKα radiation. Complete datasets (Table 1) were then collected to 2.5 Å resolution from native crystals using the imaging plate systems at the EMBL outstation at DESY and from the mercury and platinum derivatives on film at SRS Daresbury. The electron density map (Fig. 1) at 2.5 Å resolution calculated with phases from multiple isomorphous replacement (mean figure of merit, 0.63) was easily interpretable and was improved by solvent flattening[21,22]. A continuous polypeptide chain from residue 61 to residue 644 at the C terminus was traced unambiguously, and most side-chain atoms could be located in the map. The atomic model was built using the graphics program O (ref. 23) and had an initial R-factor of 37% for all data to 2.5 Å. After preliminary refinement using the program X-PLOR (ref. 24), the current model, containing 584 amino acid residues and 40 bound water molecules, has an R-factor of 19.9% and r.m.s. bond length deviation of 0.017 Å.

Description of the structure

Overview. The beetle toxin is a wedge-shaped molecule with a radius of gyration of 58 Å. As shown in Fig. 2*a*, it comprises three domains. Domain I, from the N terminus of the 67K toxin to residue 290, is a seven-helix bundle in which a central helix is completely surrounded by six outer helices tilted at about +20° to it (Fig. 3*b*,*c*). Domain II, from residues 291 to 500, contains three antiparallel β sheets packed around a hydrophobic core with a triangular cross-section (Fig. 4). Domain III, from residues 501 to 644 at the C terminus is a sandwich of two antiparallel β sheets (Fig. 5). Domains I and III make up the

TABLE 1 Data collection and phasing statistics

Data collection

Data	Method of collection	Number of crystals	Resolution (Å)	Number of measurements	Unique reflections (% completeness)	R_{merge}
Native	image plate	8	2.5	121,767	27,727 (100)	0.108
CH_3HgNO_3	film	7	2.5	103,623	27,767 (100)	0.095
$Hg(CH_3COO)_2$	film	5	2.5	60,224	25,919 (94.5)	0.103
cis-$Pt(NH_3)_2Cl_2$	film	7	2.5	86,629	25,924 (94.5)	0.107
K_2OsO_4	FAST	1	4.5	21,143	4,680 (100)	0.077
$HoCl_3$	FAST	1	4.5	20,013	4,701 (100)	0.069

Phasing statistics

Derivative	Anomalous data	Number of sites	R_{deriv}†	R_{Cullis}‡	Phasing power§ (resolution, Å)
CH_3HgNO_3	no	3	0.183	0.715	1.56 (2.5)
$Hg(CH_3COO)_2$	yes	6	0.247	0.609	2.28 (2.5)
cis-$Pt(NH_3)_2Cl_2$	no	5	0.185	0.682	1.54 (2.5)
K_2OsO_4	no	4	0.149	0.757	1.26 (5.5)
$HoCl_3$	no	3	0.095	0.741	1.35 (5.0)

Protein preparation: Solubilized parasporal crystals from *B.t.* subsp *tenebrionis* were incubated at 0.5 mg ml^{-1} protein with 0.125 units per ml of Agarose-linked papain (Boehringer) in 3.3 M NaBr, 0.05 M sodium phosphate, pH 7.0, and 0.1 mg ml^{-1} phenylmethylsulphonylfluoride (PMSF) for 30 min at 20 °C. Digestion was stopped by adding tosyl lysinechloromethylketone (TLCK) to 0.125 mg m^{-1} and Na_2CO_3 to one fifth volume and removing the enzyme-beads. The 67K beetle toxin was then purified by gel filtration on Sephadex G75 equilibrated with 0.1 M $NaHCO_3$, pH 10.5, 0.5 M NaBr. **Crystallization:** Single crystals were obtained by microdialysis at a protein concentration of 2.5 mg ml^{-1} against 0.1 M $NaHCO_3$, pH 9.5, 1.2 M NaBr at 4 °C overnight, then against 0.1 M $NaHCO_3$, pH 9.2, 0.5 M NaBr at 16 °C; 3 mM NaN_3, 0.1 mM PMSF and 0.1 mg ml^{-1} TLCK were present in all buffers. Crystals were transferred by stages to 0.05 M 2-(*N*-morpholino)ethanesulphonic acid (MES), pH 6.5, for derivative preparation and mounted in 0.03% low-melting agarose in this buffer during data collection. **Data collection:** Image plate and film data were processed using MOSFLM (Imperial College, London) and CCP4 programs (Daresbury, UK). FAST (ref. 20) data were collected and processed with MADNES[45], and scaled in 3° batches. **Derivatives:** Crystals were soaked respectively in 0.25 mM CH_3HgNO_3 for 3.5 h, in 1 mM $Hg(CH_3COO)_2$ for 14 h, in freshly prepared 1 mM *cis*-$Pt(NH_3)_2Cl_2$ for 21 h, in saturated K_2OsO_4 for 35 h, and in 2 mM $HoCl_3$ for 3 days. **Phase calculation:** Two heavy-atom sites in each derivative were located from difference Patterson functions, except in the case of $Hg(CH_3COO)_2$ for which 3 sites were located, and the remaining sites were found by cross-phased difference Fouriers. Heavy-atom parameters were refined against centric data and phases calculated for all data using the program PHARE (G. Bricogne). The two low-resolution derivatives were refined against phases calculated from the high-resolution derivatives. Phasing with the three high-resolution derivatives gave an overall figure of merit of 0.61 (25–2.5 Å) and a clearly interpretable map. Including the remaining derivatives slightly improved the connectivity of the map (overall figure of merit 0.63), and four cycles of solvent flattening using a 50% solvent content and a 9 Å radius in mask calculation[21,22] improved the overall definition of densities. The starting model was built using the program *O* (ref. 23) with the Bones option for main-chain tracing and the autobuild and manip options for side chains. Refinement by simulated annealing using the program X-PLOR (ref. 24) reduced the *R*-factor from 0.37 to 0.25 without individual *B*-factors, and to 0.23 with restrained individual *B*-factors. The model was adjusted in the loops 154–156, 429–436, and 483–488, and had 40 solvent molecules added, then refined by X-PLOR again. The current model has an *R*-factor of 19.9%, with r.m.s. bond length deviation of 0.017 Å, r.m.s. bond angle deviation of 3.2°, and average atomic *B*-factor of 18 Å^2.

* $R_{merge} = \sum_i \sum |I_i - \langle I \rangle| / \sum |\langle I \rangle|$, where I_i are intensity measurements for a reflection, and $\langle I \rangle$ is the mean intensity for this reflection.

† $R_{deriv} = \sum |F_{PH} - F_P| / \sum |F_P|$, where F_{PH} is the structure factor amplitude of the derivative crystal and F_P is that of the native.

‡ $R_{Cullis} = \sum ||F_{PH} \pm F_P| - F_H(calc)| / \sum |F_{PH} - F_P|$, where F_P and F_{PH} are defined as for R_{deriv}, and $F_H(calc)$ is the calculated heavy-atom structure factor amplitude summed over centric data only.

§ Phasing power = $\langle F_H \rangle / E$, the r.m.s. heavy-atom structure factor amplitudes divided by the residual lack of closure error.

bulky end of the molecule. Through their contact one of the two β sheets in domain III is almost entirely buried. To our knowledge (see, for example, ref. 25), the packing of helices in domain I and of sheets in domain II are both novel arrangements.

Domain I. The central helix in this seven-helix bundle is α_5 (Fig. 3*b*,*c*), which is oriented with its C terminus towards the bulky end of the molecule. Viewed from this end, the outer helices are arranged anticlockwise in the order of α_1, α_2, α_3, α_4, α_6 and α_7, with helices α_1 and α_7 adjacent to the β-sheet domains; α_2 is interrupted by a non-helical section and only the leading half, α_{2a}, is packed against α_5. Figure 3*a* shows the alignment of amino-acid sequence on the surfaces of the helices. The helices are long, especially α_3 to α_7, which contain respectively 8, 7, 6, 9 and 7 complete helical turns and hence would be long enough to span the 30-Å thick hydrophobic region of a membrane bilayer. Furthermore, the six outer helices bear a strip of hydrophobic residues (defined by $\Delta G \geq 0$ for transfer from oil to water) down their entire length on the side-facing helix α_5, so they are amphipathic. In keeping with the general observation that secondary structures are close-packed and bury hydrophobic surfaces[26], the helix contact angles in this domain cluster around +20° rather than −50°, giving the bundle a bouquet-like appearance (Fig. 3*b*). Figure 3*c* shows the bundle in cross-section. The interhelical space contains 27 aromatic residues which are packed in the edge-to-face fashion[27]; all polar groups in this region are hydrogen-bonded or in salt bridges.

The concentric arrangement of the seven-helix bundle is distinct from the two-layered type seen in bacteriorhodopsin. There is some resemblance to the pore-forming domain of colicin A[28], in which two hydrophobic helices are shielded from solvent by eight amphiphilic helices, but the colicin helices are generally shorter. Like the colicin helices, the bundle in the beetle toxin may be a soluble form of packaging for the hydrophobic and amphiphilic helices that will form pores in the membrane after a large change in conformation.

Domain II. In Fig. 4*a* and 4*b* the three sheets of this domain are laid side-by-side, as they would be seen from the solvent. There is an apparent structural duplication between the four-stranded antiparallel sheets, sheet 1 and sheet 2. The chain connections, β_4, β_3, β_2, β_5 and β_8, β_7, β_6, β_9, respectively, follow the order of +3, −1, −1, +3, which is typical of the 'Greek-key' topology[29]. From both sheets the inner strands, β_3 and β_2 as well as β_7 and β_6, extend some 20 Å to the apex of the molecule as two-stranded β ribbons; and at the point of departure from the sheets there is a β-bulge in β_3 and in β_7 to twist the plane of the ribbon by nearly 90° relative to the sheet. The connections between the outer strands cross over the ribbons on the solvent side.

The pseudo-symmetry between these sheets is very approximate. Using the least squares option in *O* (ref. 23), the sheet region of the strands β_3 and β_2 can be brought to superimpose on that of β_7 and β_6, with a r.m.s. fit of 0.72 Å for 13 α carbons. But the r.m.s. fit increased to 1.1 Å for 23 α carbons of the

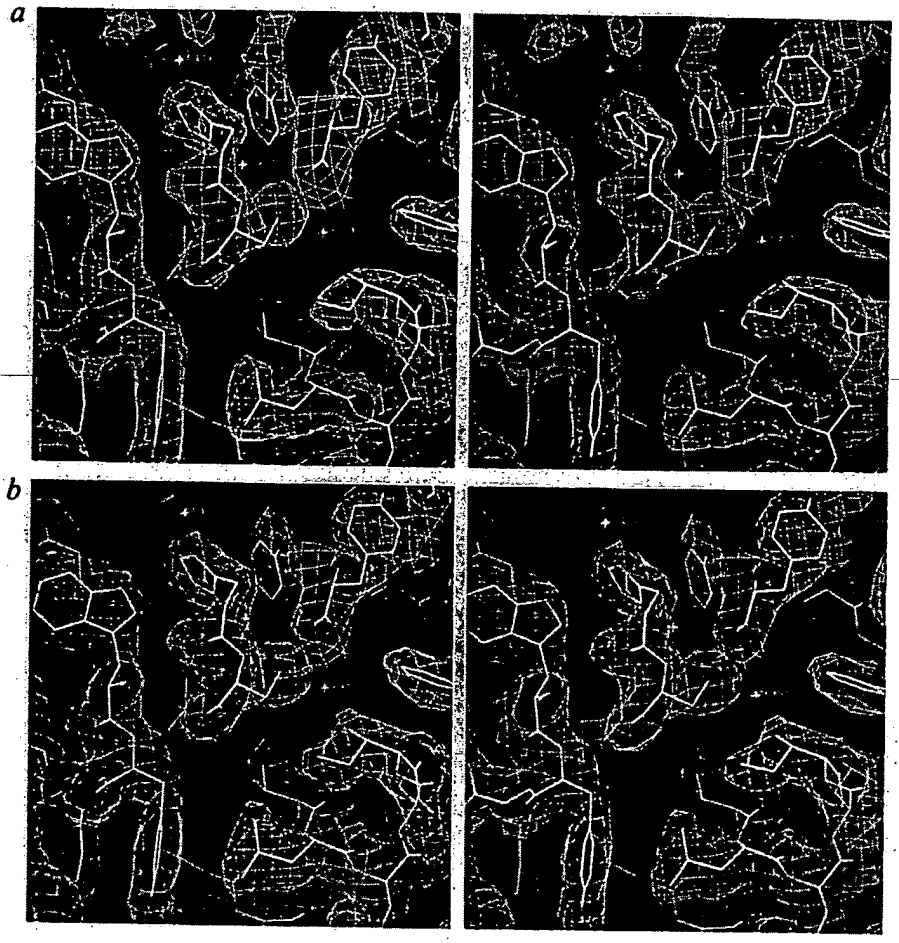


FIG. 1 Electron density map in the neighbourhood of Cys 243, calculated *a*, using combined phases[46] from multiple isomorphous replacement and solvent flattening, and *b*, using combined experimental and model phases[46] after refinement by X-PLOR. The refined structure is shown superimposed for reference. Although Cys 243 is a major site of both the methylmercury (MM) and mercuric acetate (MA) derivatives, the methyl mercury site is in a hydrophobic environment compared with the mercuric acetate site.

whole inner strands including the ribbon region, and 1.7 Å for 36 α carbons on all four strands. Nonetheless, the sequence alignment brought by this superposition of the two sheets revealed a low level of internal homology, with seven pairs of equivalent residues (shown in bold) out of 41 aligned α carbons:

```
338  HRIQFHTRFQP(6)SFNYWS(1)NYVSTRPSI(0)GSNDIITSPF(10)NLKFN  395
402  AVANTNLAVWP(0)SAVYSG(1)TKVEFSQYN(3)DEASTQTYDS(7 )SWDSI  453
```

The three-stranded sheet 3 is formed by two separate polypeptide segments. The C-terminal segment of domain II contributes the two-stranded ribbon of β_{10} and β_{11}, whereas the N-terminal segment of this domain contributes strand β_1, which is hydrogen-bonded to β_{11}; β_1 is followed by a two-turn helix α_8 and an extended chain.

Figure 4*c* and *d* shows in side view and in cross-section that the three antiparallel sheets are packed around a triangular hydrophobic core. This brings the strand β_{10} on the edge of sheet 3 into proximity with strand β_4 on the edge of sheet 1, as well as placing the loops at the end of the three β ribbons into a region of about 12 Å radius at the molecular apex. This domain is in contact with helix α_7 of domain I on the face of sheet 3 (Fig. 4*c*).

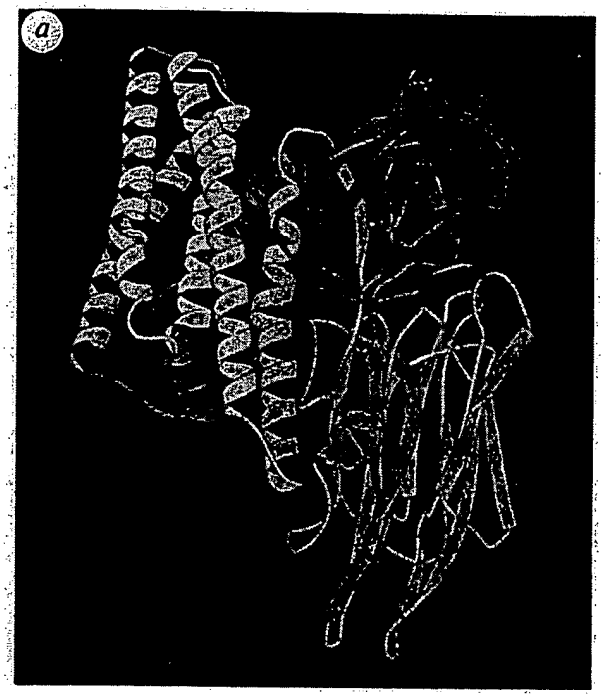
Domain III. Figure 5 is a ribbon drawing of the strands forming the two sheets of the β sandwich. The sheet containing the C-terminal strand is in contact with domain I and will be called the inner sheet. This domain has the 'jelly-roll' topology[29], because it can be generated by folding an antiparallel β ribbon which starts with β_{13} (N terminus) and β_{23} (C terminus) on the inner sheet, and ends in the loop between β_{18} and β_{19} on the outer sheet; β_{14} is a short excursion from this ribbon and forms the fifth antiparallel strand of the outer sheet. In addition, small parallel sheets are formed at the edge of the β sandwich through hydrogen bonding of strand β_{12} to β_{16} at the edge of the outer sheet, and β_1 to β_{13} at the edge of the inner sheet.

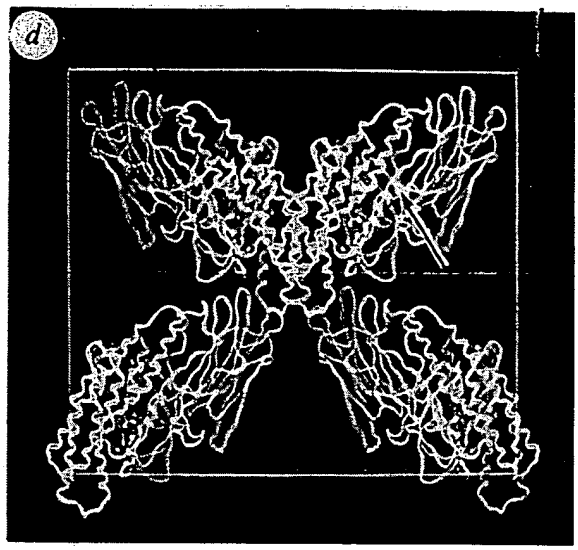
Distribution of conserved sequences. The core of the beetle toxin molecule encompassing the domain interfaces is built from the five sequence blocks that are highly conserved throughout the δ-endotoxin family[1] (Fig. 2*b*,*c*). Block 1, located in the beetle toxin sequence at residues 189–218, corresponds to the central helix (α_5) of the bundle in domain I. Block 2, residues 239–305, overlaps with the latter half of α_6, and with α_7 and β_1; the latter hydrogen-bonds to the edge of the inner sheet in domain III before forming part of the three-stranded sheet 3 in domain II. Block 3, residues 491–538, overlaps with the latter part of β_{11}, where it is hydrogen-bonded to β_1, and with the loops connecting domains II and III. The remainder of block 3 together with blocks 4 and 5, namely residues 560–569 and 633 to the C terminus, respectively, constitute the three buried strands of the inner antiparallel sheet in domain III. The high degree of conservation of internal residues implies that homologous proteins would adopt a similar fold. Using the beetle toxin structure as a model, we can therefore propose a basis for the insecticidal activity of δ-endotoxins as a family.

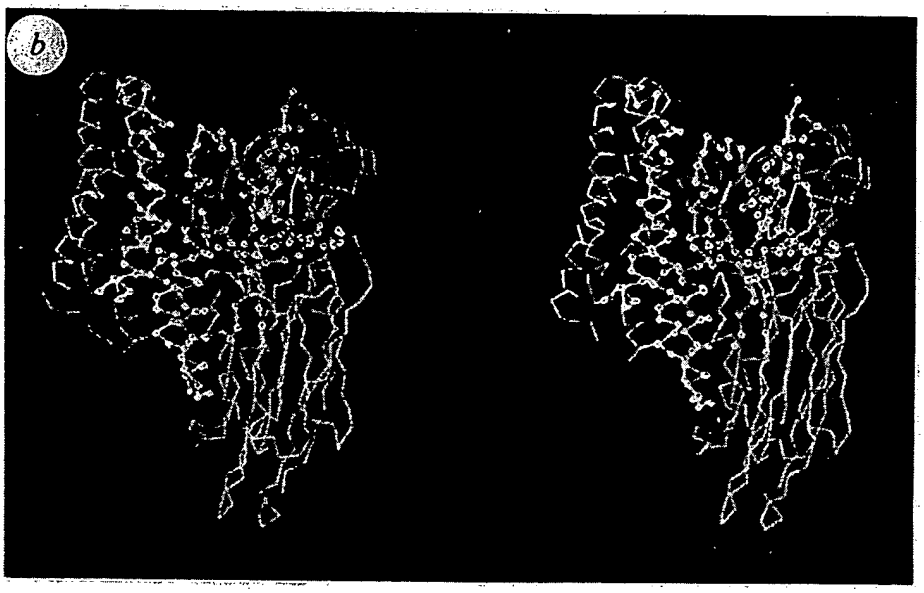
Basis of insecticidal function

Solubility. The beetle toxin crystals are isomorphous with the parasporal crystals[18,19] and show the molecular contacts responsible for solubility behaviour *in vivo*. Four intermolecular salt bridges, Asp 142–Arg 165, Asp 224–Arg 562, Asp 590–Arg 178, and Glu 223–Lys 293, are located at contacts to three different neighbouring molecules. Such salt bridges keep the protoxin crystals insoluble until exposed to the extreme pHs in the insect midgut.

Proteolytic activation. Pro-δ-endotoxins have M_rs of either ~130K or ~70K. Activation by larval gut proteases removes the C-terminal half of the larger protoxins[30,31] and cleaves them at residue 28 or 29 from the N terminus. The smaller protoxins, such as that of the beetle toxin, are processed only at the N







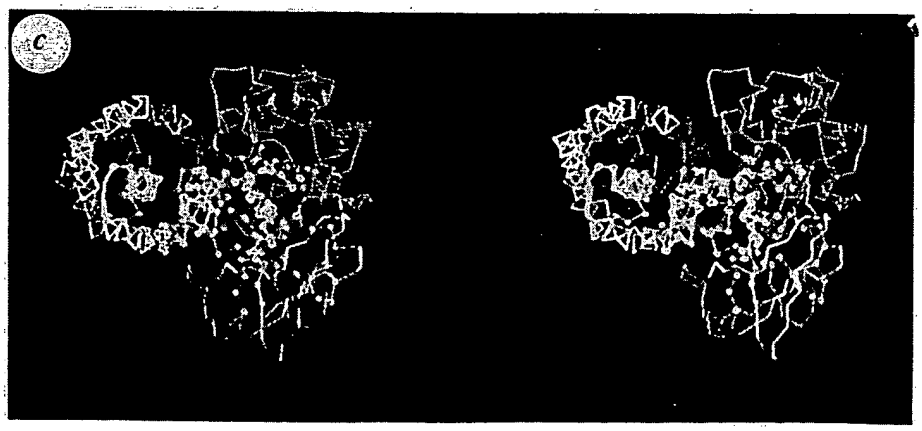


FIG. 2 Overview. *a*, Schematic ribbon representation of the beetle toxin showing the domain organization. Secondary structure assignments are given by Yasspa within program *O* (ref. 23). The polypeptide pathway is indicated by colouring the chain in the rainbow order, from red at the N terminus to blue at the C terminus. The three domains are: I, a seven-helix bundle (upper left); II, a three-sheet assembly (bottom); and III, a β sandwich (upper right). This and all following illustrations of the structure are made with the program MOLSCRIPT[47]. *b* and *c*, Cα trace (stereoview) of the molecule with the five conserved sequence blocks indicated by small beads at their Cα positions. In *b* the view is as in *a* and in *c* it is down the central helix of the bundle from the bulky end of the molecule; *c* shows that the central helix of domain I and the inner sheet of domain III are conserved; *b* shows that the helices at the domain I–II interface and the loops at the domain II–III interface are also conserved. Note in *c* the helix packing of six around one in domain I. *d*, The solvent channel in the $C222_1$ lattice viewed along the *c* axis. One half of the unit cell thickness is shown, containing four molecules. The other half of the cell is related to this by a two-fold rotation about horizontal axes (blue lines) at $(\frac{1}{2}, y, \pm\frac{1}{4})$. The stacking of both layers leaves solvent channels that traverse the cell along the *c* direction. The N terminus of the molecule (arrow) is accessible from these channels.

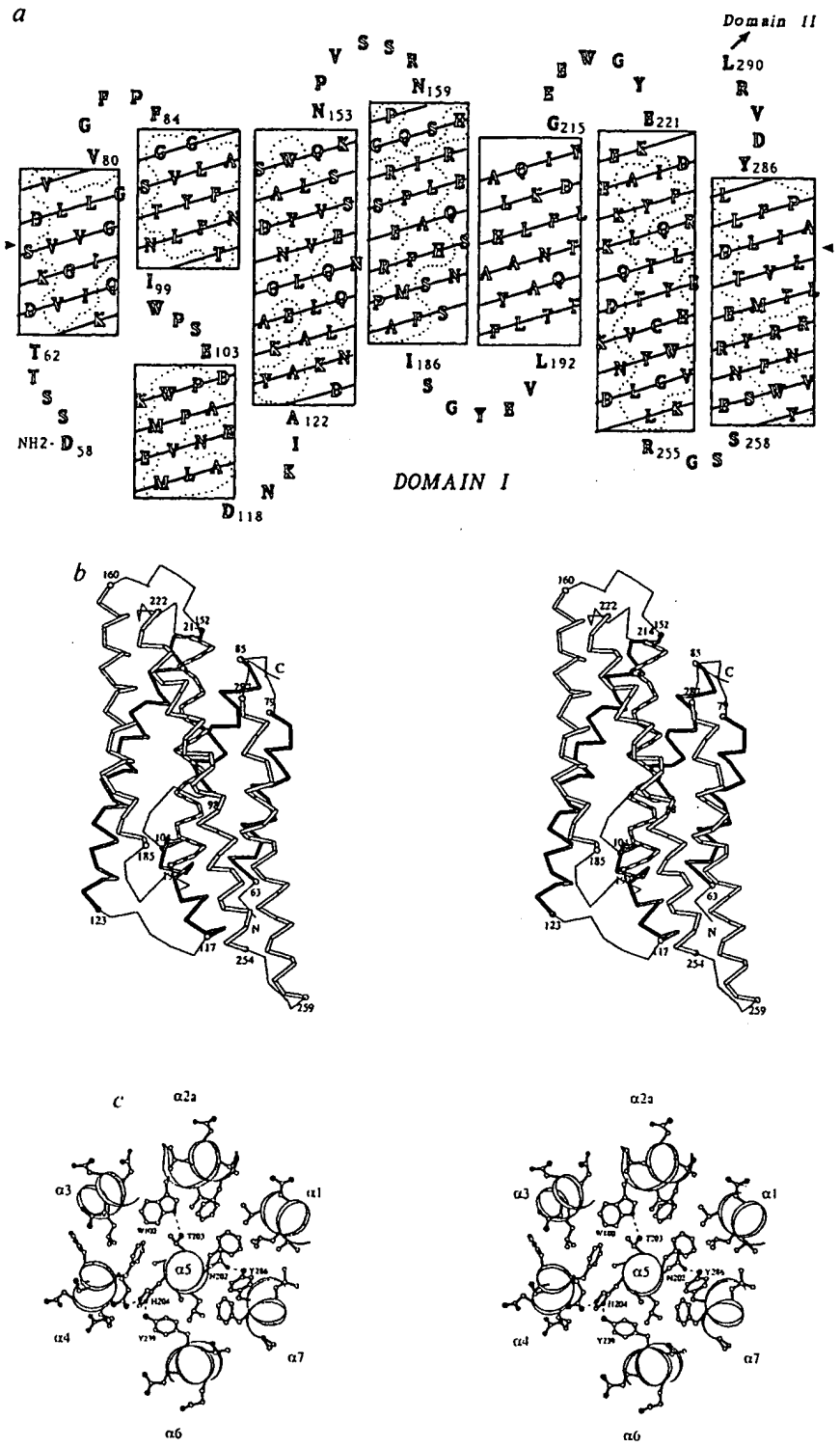


FIG. 3 The seven-helix bundle. *a*, Helical nets showing the position of amino-acid residues along the 7 helices: α_1 (63-79); α_2 (α_{2a}, 85-98 and α_{2b}, 104-117), σ_3 (123-152), α_4 (160-185), α_5 (193-214), α_6 (222-254) and α_7 (259-285). The cylindrical surface of the helices are cut longitudinally on the side facing the solvent and flattened to give a view from the interior of the bundle. The top of the drawing corresponds to the bulky end of the whole molecule. Owing to tilting of the outer helices, different helices are in register vertically only at a level indicated by two arrows pointed at α_1 and α_7; α_5 is the central helix. Dotted curves outline the strip of hydrophobic residues down the inward surface of the other six helices. *b*, Cα trace (stereoview) for the bundle viewed perpendicular to α_5. The relative tilt of the outer helices to α_5 and that between adjacent outer heleices are both about 20°. The Cα trace is shaded grey over helices α1 to α3 in the back, striped over helix α5 in the centre, and white over helices α4, α5, and α7 in the front. *c*, Cross-section of the bundle at the level indicated by the arrows in *a*, viewed from the bulky end of the molecule. The hellical backbone is represented by curly ribbons passing through the Cα positions. The outer helices are positioned roughly hexagonally around the central one and tilted relative to it, so the bundle forms a left-handed superhelix. The aromatic side chains are packed in an edge-to-face fashion. Hydrogen bonds are shown for side-chain atoms.

terminus[19,32] where about 50 residues are removed. The activated δ-endotoxins show a conserved C-terminus, so-called sequence block 5 (ref. 1). Its position as the middle strand of the buried β sheet in domain III precludes further processing from the C terminus. In fact deletion from this site by 4 to 8 residues results in inactive mutants with altered solubility and immunogenicity[30,33-35]. This is not surprising as the inner sheet can be expected to play a critical part in the structural integrity and stability of the toxins through interaction with the helical bundle.

At the N-terminal cleavage sites the different protoxin sequences show locally similar hydropathy profiles[36,37], which would be consistent with a common topology for the N-terminal region of the activated toxins as seen in the helical bundle of the beetle toxin. In crystals of the beetle toxin, the N terminus at the start of helix α_1 borders on a large solvent channel of about 30 Å diameter that crosses the unit cell along the *c* direction (Fig. 2*d*). This channel could allow access of sporulation-associated proteases to the cleavage site in parasporal crystals[19].

Receptor binding. The insecticidal selectivity of δ-endotoxins is due to high-affinity binding to specific membrane receptors[7-9,38], which in three cases seem to be glycoproteins[38-40]. For several δ-endotoxins the specificity-determining regions have been delimited by exchanging sequence segments between closely related toxins of differing specificities[13-15]. Guided by the location of secondary structures in the beetle toxin, a plausible alignment of δ-endotoxin sequences was made for the non-

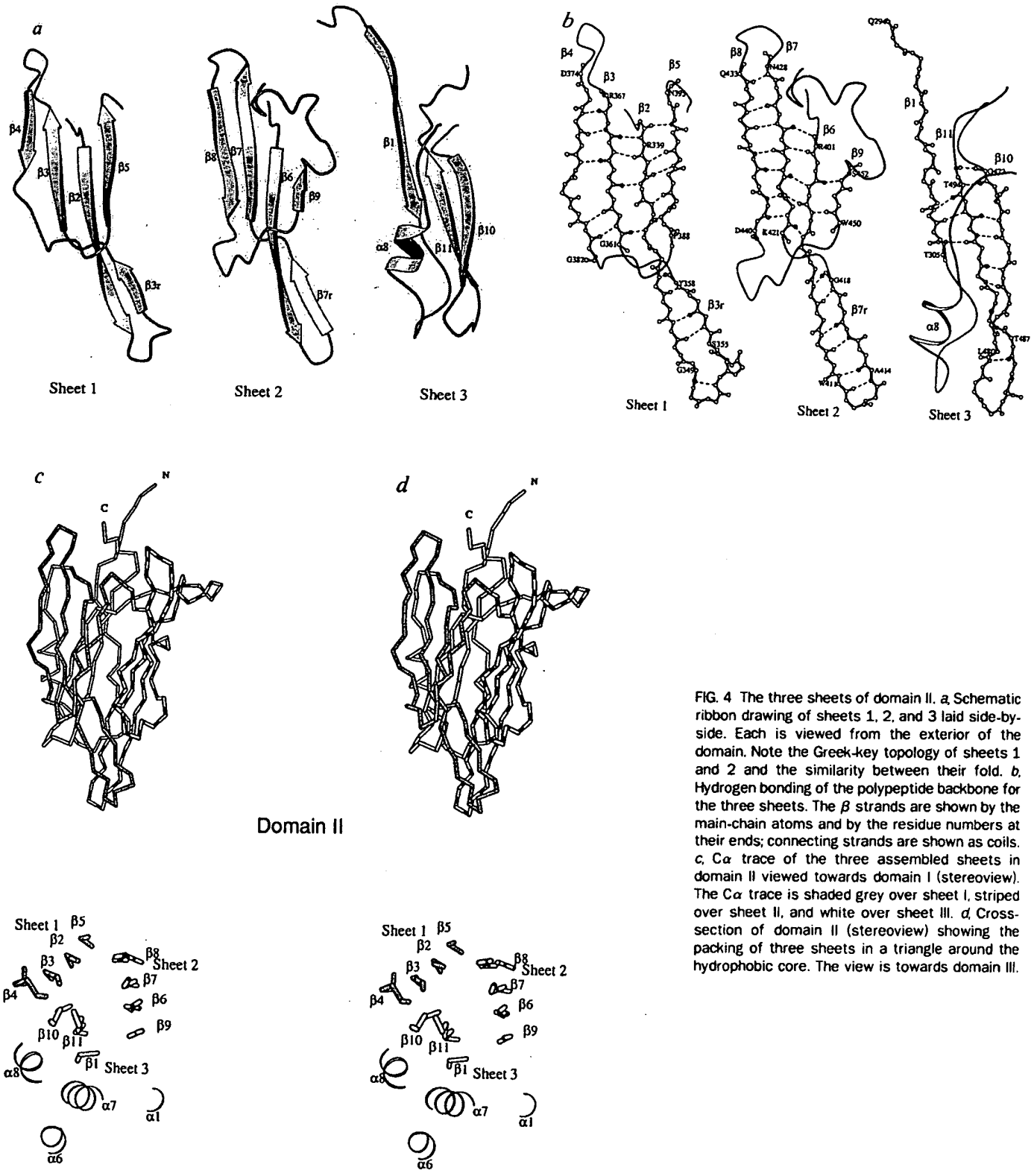
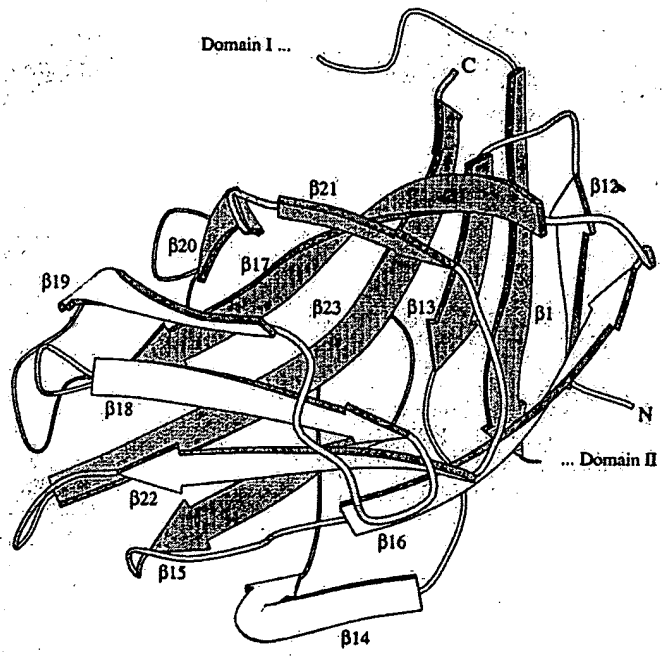


FIG. 4 The three sheets of domain II. *a*, Schematic ribbon drawing of sheets 1, 2, and 3 laid side-by-side. Each is viewed from the exterior of the domain. Note the Greek-key topology of sheets 1 and 2 and the similarity between their fold. *b*, Hydrogen bonding of the polypeptide backbone for the three sheets. The β strands are shown by the main-chain atoms and by the residue numbers at their ends; connecting strands are shown as coils. *c*, Cα trace of the three assembled sheets in domain II viewed towards domain I (stereoview). The Cα trace is shaded grey over sheet I, striped over sheet II, and white over sheet III. *d*, Cross-section of domain II (stereoview) showing the packing of three sheets in a triangle around the hydrophobic core. The view is towards domain III.

conserved regions (ref. 12, and T. C. Hodgman, unpublished results). Hence the genetically identified specificity-determining regions can be mapped to equivalent positions in the beetle toxin structure, and these fall mainly in domain II. For instance, the dual specificity of CryIIA for Lepidoptera and Diptera, as distinct from the Lepidoptera specificity in the closely related CryIIB, is determined by residues 307–382 of their sequences[14], which corresponds roughly to sheet 1 (Fig. 4*a*) plus strand β_6 in sheet 2 and the loop leading up to β_7, whereas the Lepidoptera specificity of CryIIB is dependent on a longer segment[14] that would include both inner strands of sheet 2. Similarly, the toxicities of CryIA(a) and CryIA(c) to two lepidopteran insects depend on three segments termed x, y and z (ref. 15): amino-acid substitutions in y can reduce toxicity by up to 2,000-fold, and segments x and y interact in determining specificity. Aligned with the beetle toxin structure, segment x corresponds roughly to the outer strands β_4 and β_5 of sheet 1 and the whole of sheet 2, including the loop entering β_{10} in sheet 3; y corresponds to



Domain III

FIG. 5 Domain III, schematic ribbon representation of the β sandwich. β strands forming the inner sheet are shaded grey. The topology of an eight-stranded 'jelly-roll' can be seen by following the β hairpin starting with β_{13}, β_{15} and β_{23} in the inner sheet, continuing to β_{16} and β_{22} in the outer sheet, then β_{17} and β_{21}, β_{20} in the inner sheet, and ending with β_{18} and β_{19} in the outer sheet. β_{14} is an excursion from the hairpin and forms a fifth antiparallel strand of the outer sheet. Small parallel β sheets are added to one edge of the β sandwich, by hydrogen bonding of β_1 to β_{13} in the inner sheet and β_{12} to β_{16} in the outer sheet. Residue numbers in the β strands are: β_{12}, 502-506; β_{13}, 509-513; β_{14}, 519-525; β_{15}, 536-541; β_{16}, 547-554; β_{17}, 558-569; β_{18}, 573-579; β_{19}, 585-591; β_{20}, 604-609; β_{21}, 611-614; β_{22}, 619-625; and β_{23}, 631-643.

strand β_{10} of sheet 3 and the loop connecting β_{10} and β_{11}; and z extends from β_{11} to the C-terminal activation site. Furthermore, the interaction between x and y can be understood in terms of the proximity between β_4 on the edge of sheet 1 and β_{10} on the edge of sheet 3. Although z was inferred[15] to extend into domain III, the combined evidence from genetics and receptor-binding assays *in vitro* for Lepidoptera toxins[9,41] correlates receptor recognition with sequence variations within domain II. We note that the β ribbons from all three sheets terminate in loops in a small region on the molecular apex, in a manner reminiscent of the complementarity-determining region of immunoglobins.

Pore formation. The common mechanism of epithelial cell disruption by δ-endotoxins of widely different specificities is believed to be the formation of lytic pores of 10 to 20 Å diameter in the insect membrane[10]. The structure of the beetle toxin displays an apparatus for pore formation in the long, hydrophobic and amphipathic helices of domain I which could penetrate the membrane. Between the crystal structure in which the bouquet-like helical bundle internalizes all the hydrophobic surfaces, and the unknown pore structure where hydrophobic surfaces would be in intimate contact with the membrane lipids, large conformation changes must occur. In the absence of a full characterization of the pore-forming process, we propose the following by extrapolation from the crystal structure.

The trigger for the conformational changes may be provided by receptor binding and the consequent interaction of toxin with the membrane bilayer. Membrane insertion follows rapidly, so that a major part of the bound δ-endotoxin cannot be displaced from the brush-border vesicles by other toxins recognizing the same receptor sites[7,9]. As domain II and probably its apical region are most likely to bind the membrane receptors, the helices are expected to insert with the 'domain II end' (see Fig. 2*a*) oriented towards the cytoplasm. If helical hairpins are to initiate the membrane penetration, as probably happens for colicin[28,42,43], they will probably be linked at the domain II end. So either of the helix pairs α_6-α_7 or α_4-α_5 could be the likely initiator. The α_6-α_7 pair is favoured because it forms part of the conserved interface with domain II and is well positioned to sense the receptor binding. On the other hand, helix α_5 is the most conserved throughout the family of δ-endotoxins. Point mutations in α_5 reduce toxicity of a Lepidoptera toxin without reducing binding to membranes[44]. Proteolysis in the interhelical loops at the domain III end, as in the α_3-α_4 loop[19,32], may facilitate release of the helix pairs from the tertiary structure of the bundle. The insertion of a hairpin can create a defect in the membrane, allowing the rest of domain I to participate in pore formation in a cooperative manner. □

Received 22 July; accepted 19 September 1991.

1. Höfte, H. & Whiteley, H. R., *Microbiol. Rev.* **53**, 242-255 (1989).
2. Ellar, D. J. *et al.* in *Molecular Biology of Microbial Differentiation* (eds Hoch, J. A. & Setlow, P.) 230-240 (Am. Soc. Microbiol., Washington, DC, 1985).
3. Wilcox, E. R. *et al.* in *Protein Engineering: Applications in Science, Medicine and Industry* (eds Inouye, M. & Sarma, R.) 395-413 (Academic, New York, 1986).
4. Vaeck, M. *et al. Nature* **328**, 33-37 (1987).
5. Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, D. L. & Fischhoff, D. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3324-3328 (1991).
6. Barton, K. A., Whiteley, H. R. & Yang, N. S. *Plant Physiol.* **85**, 1103-1109 (1987).
7. Hofmann, C., Lüthy, P., Hütter, R. & Pliska, V. *Eur. J. Biochem.* **173**, 85-91 (1988).
8. Hofmann, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7844-7848 (1988).
9. Van Rie, J., Jansens, S., Höfte, H., Degheele, D. & Van Mellaert, H. *Appl. Envir. Microbiol.* **56**, 1378-1385 (1990).
10. Knowles, B. H. & Ellar, D. J. *Biochim. biophys. Acta* **924**, 509-518 (1987).
11. Endo, Y. & Nishitsutsuji-Uwo, J. *J. Invertebr. Path.* **36**, 90-103 (1980).
12. Hodgman, T. C. & Ellar, D. J. *J. DNA Sequ. Map.* **1**, 97-106 (1990).
13. Ge, A. Z., Shivarova, N. I. & Dean, D. H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4037-4041 (1989).
14. Widner, W. R. & Whiteley, H. R. *J. Bact.* **172**, 2826-2832 (1990).
15. Schnepf, H. E., Tomczak, K., Ortega, J. P. & Whiteley, H. R. *J. biol. Chem.* **265**, 20923-20939 (1990).
16. Krieg, A., Huger, A. M., Langenbruch, G. A. & Schnetter, W. *J. appl. Entomol.* **96**, 500-508 (1983).
17. Höfte, H., Seurinck, J., Van Houtven, A. & Vaeck, M. *Nucleic Acids Res.* **15**, 7183 (1987).
18. Li, J., Henderson, R., Carroll, J. & Ellar, D. *J. molec. Biol.* **199**, 543-545 (1988).
19. Carroll, J., Li, J. & Ellar, D. J. *Biochem. J.* **261**, 99-105 (1989).
20. Arndt, U. W. *Meth. Enzym.* **114**, 472-485 (1985).
21. Wang, B. C. *Meth. Enzym.* **115**, 90-112 (1985).
22. Leslie, A. G. W. *Acta crystallogr.* **A43**, 134-136 (1987).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, N. *Acta crystallogr.* **A47**, 110-119 (1991).
24. Brünger, A. T. *J. molec. Biol.* **203**, 803-816 (1988).
25. Chothia, C. & Finkelstein, A. V. *A. Rev. Biochem.* **59**, 1007-1039 (1990).
26. Chothia, C. *J. molec. Biol.* **105**, 1-15 (1976).
27. Burley, S. K. & Petsko, G. A. *Science* **229**, 23-28 (1985).
28. Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. *Nature* **337**, 93-96 (1989).
29. Richardson, J. S. *Adv. Prot. Chem.* **34**, 167-339 (1981).
30. Höfte, H. *et al. Eur. J. Biochem.* **161**, 273-280 (1986).
31. Choma, C. T. & Kaplan, H. *Biochemistry* **29**, 10971-10977 (1990).
32. Nicholls, C. N., Ahmad, W. & Ellar, D. J. *J. Bact.* **171**, 5141-5147 (1989).
33. MacIntosh, S. C., McPherson, S. L., Perlak, F. J., Marrone, P. G. & Fuchs, R. L. *Biochem. biophys. Res. Commun.* **170**, 665-672 (1990).
34. Schnepf, H. E. & Whiteley, H. R. *J. biol. Chem.* **260**, 6273-6280 (1985).
35. Adang, M. J. *et al. Gene* **36**, 289-300 (1985).
36. Sekar, R., Thompson, D. V., Maroney, M. J., Bookland, R. G. & Adang, M. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7036-7040 (1987).
37. Chungjatupornchai, W., Höfte, H., Seurinck, J., Angsuthanasombat, C. & Vaeck, M. *Eur. J. Biochem.* **173**, 9-16 (1988).
38. Knowles, B. H., Thomas, W. E. & Ellar, D. J. *FEBS Lett.* **168**, 197-202 (1984).
39. Knowles, B. H. & Ellar, D. J. *J. Cell Sci.* **83**, 89-101 (1986).
40. Haider, M. Z. & Ellar, D. J. *Biochem. J.* **248**, 197-201 (1987).
41. Visser, B., Munsterman, E., Stoker, A. & Dirske, W. G. *J. Bact.* **172**, 6783-6788 (1990).
42. Lakey, J. H., Baty, D. & Pattus, F. *J. molec. Biol.* **218**, 639-653 (1991).
43. Song, H. Y., Cohen, F. S. & Cramer, W. A. *J. Bact.* **173**, 2927-2934 (1991).
44. Ahmad, W. & Ellar, D. J. *FEMS Microbiol. Lett.* **68**, 97-104 (1990).
45. Messerschimidt, A. & Pflugrath, J. W. *J. appl. Crystallogr.* **20**, 306-315 (1987).
46. Read, R. J. *Acta crystallogr.* **A42**, 140-149 (1986).
47. Kraulis, P. *J. appl. Crystallogr.* (in the press).

ACKNOWLEDGEMENTS. We thank P. R. Evans, A. G. W. Leslie and R. Henderson for advice and encouragement; K. Wilson and Z. Dauter for help with the image plate system; K. Nagai and P. J. McLaughlin for help in collecting film data; T. A. Jones for advice on model building; SERC Daresbury Laboratory and EMBL Outstation at DESY for use of synchrotron facilities; and T. Woollard and K. Hopkins for maintaining the rotating anodes. D.J.E. and J.C. acknowledge the support of the AFRC.

Structure, Vol. 9, 409–417, May 1, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00601-3

Structure of Cry2Aa Suggests an Unexpected Receptor Binding Epitope

R.J. Morse,[1] T. Yamamoto,[2,4] and R.M. Stroud[1,3]
[1]Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143
[2]Sandoz Agro
975 California Avenue
Palo Alto, California 94304

Summary

Background: Genetically modified (GM) crops that express insecticidal protein toxins are an integral part of modern agriculture. Proteins produced by *Bacillus thuringiensis* (Bt) during sporulation mediate the pathogenicity of Bt toward a spectrum of insect larvae whose breadth depends upon the Bt strain. These transmembrane channel-forming toxins are stored in Bt as crystalline inclusions called Cry proteins. These proteins are the active agents used in the majority of biorational pesticides and insect-resistant transgenic crops. Though Bt toxins are promising as a crop protection alternative and are ecologically friendlier than synthetic organic pesticides, resistance to Bt toxins by insects is recognized as a potential limitation to their application.

Results: We have determined the 2.2 Å crystal structure of the Cry2Aa protoxin by multiple isomorphous replacement. This is the first crystal structure of a Cry toxin specific to Diptera (mosquitoes and flies) and the first structure of a Cry toxin with high activity against larvae from two insect orders, Lepidoptera (moths and butterflies) and Diptera. Cry2Aa also provides the first structure of the proregion of a Cry toxin that is cleaved to generate the membrane-active toxin in the larval gut.

Conclusions: The crystal structure of Cry2Aa reported here, together with chimeric-scanning and domain-swapping mutagenesis, defines the putative receptor binding epitope on the toxin and so may allow for alteration of specificity to combat resistance or to minimize collateral effects on nontarget species. The putative receptor binding epitope of Cry2Aa identified in this study differs from that inferred from previous structural studies of other Cry toxins.

Introduction

The almost 20 million hectares of GM crop fields in North America consist of crops engineered for herbicide or insect resistance. The genes that confer the latter trait come from *Bacillus thuringiensis* (Bt), a family of Gram-positive sporulating soil bacteria that produce parasporal crystals with insecticidal activity. The insecticidal activity of particular Bt isolates is directed against narrow spectra of insect larval species, usually within a single order. Bacterial toxins known as insecticidal crystal proteins (ICPs) or crystalline (Cry) proteins that are sequestered as protoxins in crystalline inclusions after sporulation mediate this species-specific pathogenicity [1]. The Cry protoxins are ingested, solubilized in the larval gut [2, 3], and activated by the removal of an amino-terminal segment and a C-terminal segment, the size of which depends on the gene or its protoxin [2, 4]. The active toxins associate with insect-specific receptors of gut epithelial cells of the target insect [5] and subsequently insert into the cell membrane [6, 7], leading to the formation of ion channels [8, 9, 10]. This results in disruption of the electrochemical balance across the basal membrane, gut paralysis, and larval death [11, 12, 13, 14]. The host cadaver serves as growth medium for vegetative cells arising from germination of the Bt endospores.

Species selectivity of Cry proteins is encoded in the binding site for the target receptor [5]. Classification of the Cry proteins is based on amino acid sequence identity [15] and is roughly correlated with the taxonomic order of susceptible insect species, spanning species of agricultural (Cry1 Lepidoptera, Cry2 Lepidoptera, and Cry3 Coleoptera) and public health (Cry2 and Cry4 Diptera) significance. The structure may help guide mutagenesis followed by screening that is directed toward the fine tuning of species selectivity in order to design insecticides that do not kill nontarget organisms such as monarch larvae [16]. It also may assist in the elucidation of the structural basis of resistance to Bt toxins and the subsequent generation of novel insecticidal toxins for use on Bt-resistant insects [17, 18].

Structure-based protein engineering of Cry toxins may direct the search for variants with broader susceptible species spectra, optimal potency, and stability properties. Cry2Aa is among an unusual subset of Cry proteins possessing broad insect species specificity by exhibiting high specific activity against two insect orders, Lepidoptera and Diptera [19, 20]. It is lethal to more lepidopteran species than the Cry1 toxins deployed against agriculturally important Lepidoptera [21] and exhibits a low level of crossresistance in Cry1A-resistant insects [22]. Also, the mode of action of Cry2Aa may be distinct from that of other Cry toxins [23]. Thus, it could serve as a platform for the design of Cry toxins with broader susceptible species spectra and minimal Cry1A-derived crossresistance in the field.

Chimeric-scanning mutagenesis experiments have identified disjoint blocks (D and L, see Results and Discussion) of the Cry2Aa sequence that separately confer specificity against dipteran (D) and lepidopteran (L) species [24, 25]. These experiments also demonstrate that maximal activity against lepidopteran species requires not only L block residues but also some of the specificity determinants of the D residue block. Further, Cry2Ab, an 87% sequence identical homolog of Cry2Aa, has

[3]Correspondence: stroud@msg.ucsf.edu
[4]Present address: Maxygen, Redwood City, California, 94063.

Key words: Bacillus thuringiensis; delta-endotoxin; Cry2Aa; binding epitope; crystal structure; X-ray

Table 1. Data Collection, MIR Phasing, and Structure Refinement Statistics for Cry2Aa

	Native	$U(NO_3)$	PtCNS	PtI_6	$NbCl_2$	Ru[1]	Hg[2]
Data Collection (1.08 Å)							
Resolution (Å)	2.2	2.6	2.5	2.6	2.6	2.5	2.5
Unit cell dimensions (Å)	a = b = 85.6 c = 163.9						
Space group	$P4_32_12$						
Number of observed reflections (σ_F – 2.5)	245,580	69,703	139,057	70,618	73,949	113,242	126,930
Number of unique reflections	31,591	17,370	20,476	17,999	17,455	19,475	20,198
Completeness (%)	99.3	89.1	99.0	92.	89.3	94.7	97.0
R_{merge} (%)	5.7	5.1	5.7	4.7	4.3	5.4	6.3
Phasing/MIR							
Resolution		2.6	2.6	2.6	4.5	2.6	5.25
Number of sites refined		5	6	7	6	5	3
Number of reflections (σ_F 3)		16,177	17,808	16,516	3,136	17,074	2,419
R_{iso} (%)		16	24	32	10	10	16
R_{cullis}		.62	.62	.59	.60	.67	.62
R_{kraut}		.13	.15	.20	.06	.08	.09
Phasing power		1.1	1.9	1.8	1.4	0.8	1.2
$<FOM>_{centric}$[3]		.36	.39	.41	.41	.30	.42
$<FOM>_{overall}(n_{phased})$[4]	.65 (18,677)						
Refinement							
Resolution (Å)	28.0–2.2						
Number of reflections (completeness %)	31,509 (93)						
R_{cryst} [σ_F = 0] (2.3–2.2 Å)	.18 (.21)						
R_{free} [5% test] (2.3–2.2 Å)	.24 (.23)						
Number of non hydrogen atoms	5,001						
Number of water molecules	514						
Rms bond distances (Å)	.005						
Rms bond angles (°)	1.2						

[1] $[Ru(NH_3)_6]Cl_3$.
[2] Para chloromercuri phenol (PCMP).
[3] Individual data set results.
[4] Final number of phased reflections.

negligible activity against dipteran species and 3- to 8-fold less activity against certain lepidopteran species [25, 26]. Hence, Cry2Aa structure and mutagenesis data provide the basis for future protein engineering of Cry toxins with modified specificity and selectivity profiles.

To understand the structural determinants of Cry toxin specificity, we determined the crystal structure of the protoxin of Cry2Aa from *Bacillus thuringiensis* subsp. *kurstaki*. The complete structure was determined by multiple isomorphous replacement and refined to 2.2 Å resolution. We have identified a candidate toxin receptor binding surface that is consistent with available chimeric-scanning mutagenesis data.

Results and Discussion

The structure of Cry2Aa from *Bacillus thuringiensis* subsp. *kurstaki* was determined by multiple isomorphous replacement using six heavy atom derivatives and was refined to 2.2 Å resolution with R_{cryst} = 18% (Table 1). The structure of the 633-amino acid protoxin contains the N-terminal 49-amino acid peptide that is cleaved upon activation and the three domains of what will become the mature toxin [27]. The structures of the three domains are surprisingly similar in overall topology (Figure 1a) to those of the activated toxins Cry3Aa [28] and Cry1Aa [29], suggesting that removal of the activation peptide serves to expose regions of the toxin rather than alter its conformation. This structural homology is also surprising since these toxins have little sequence identity to Cry2Aa (20% and 17%, respectively). In the mature toxin, the N-terminal domain (residues 1–272) is a pore-forming seven-helical bundle (Figure 1d) [1]. The second domain (residues 273–473) is a receptor binding β prism, a three-fold symmetric arrangement of β sheets, each with a Greek key fold (Figure 1e). The third domain (residues 474–633) is implicated in determining both larval receptor binding [30, 31] and pore function [32] and is a lectin-like C-terminal β sandwich (Figure 1f).

Available chimeric-scanning mutagenesis data [24, 25] define a candidate toxin-receptor binding surface on Cry2Aa that is comprised of a distribution of hydrophobic residues (Ile474–Ala477 from β12a, Val365–Leu369 from the β5-β6 loop, and Leu402–Leu404 from the β7-β8 loop) across the solvent-exposed surface of the β prism and β sandwich domains (Figure 1b). Proteolytic activation of the toxin involves the removal of the 49 N-terminal amino acids and exposes residues comprising this putative toxin-receptor binding surface. Removal of the 49 amino terminal residues, comprised of α0, α0a, and an N-terminal coil, would not affect the structure of the seven-helical membrane insertion domain, as seen by comparing the structures of the activated toxin Cry1Aa and that of the protoxin Cry2Aa.

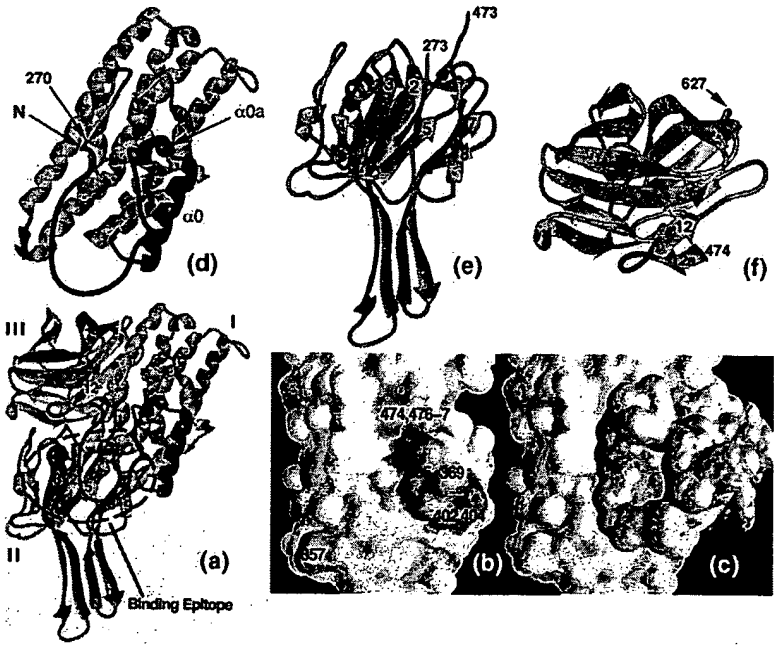


Figure 1. Topology and Solvent Accessible Surface of Cry2Aa

(a) Ribbon diagram, rendered by Midas Plus [48], of Cry2Aa. Domain I is shown in magenta, domain II is shown in blue, and domain III is shown in cyan. The N terminus is shown in red, while functionally important loops delimiting the putative toxin-receptor binding epitope are shown in green. A Cry2Aa insertion, relative to Cry3Aa and Cry1Aa, before β12 at the N terminus of domain III is shown in magenta. Numbered β strands referred to in the text are labeled.

(b) The solvent accessible surface, as calculated by GRASP [49], of domains II and III of Cry2Aa. The orientation is identical to that shown in Figure 1a. The projection of residue hydrophobicity onto this surface is shown in color. Portions of the hydrophobic surface contributed by residues 474, 476, and 477 are shown in cyan, those contributed by residues 365–369 are shown in blue, those contributed by residues 402 and 404 are shown in magenta, and the remainder of the surface contributed by hydrophobic residues is shown in yellow. The remaining surface that is identified as nonhydrophobic is colored white. Residue hydrophobicity is as defined by GRASP [49]. The prominent hydrophobic patch is the center of the putative toxin-receptor binding epitope. For orientation, the portion of the surface contributed by residue 357 of the β4-β5 loop is shown in red.

(c) The solvent accessible surface (as calculated by GRASP) of domains II and III of Cry2Aa. The orientation is identical to that shown in Figure 1a. The projection of residue hydrophobicity onto this surface is shown in yellow, while the N terminus is shown in red; the N terminus sterically hinders access to the putative toxin-receptor binding epitope. Portions of the surface that are identified as nonhydrophobic are colored white.

(d–f) The three domains of Cry2Aa shown in the same orientation as in Figure 1a. Labels with amino acid numbers identify the visible N and C termini of each domain in the figures.

This is also expected since constructs consisting of the N-terminal-helical domain of the complete Cry3Ba1 protoxin (prior to cleavage) are capable of nonreceptor-mediated partitioning into lipid bilayers [33], as is the activated toxin.

The structure of Cry2Aa suggests that the N-terminal residues should sterically hinder access to the putative binding epitope β5-β6 and β7-β8 loops (Figure 1a, shown in green) and the exposed parts of domain III closest to domain II. Projection of hydrophobicity onto the solvent accessible surface of domains II and III reveals an 800 Å² hydrophobic patch (Figure 1b) proximal to these loops. However, while the structure suggests that the 49 N-terminal residues (α0, α0a, and the N-terminal coil) should sterically hinder access to the putative binding epitope, the biological rationale for this function is unclear. It is unlikely that Bt possesses a receptor with affinity for the activated toxin. Hence, it does not seem likely that the N terminus serves to prevent premature activation of the toxin within Bt. One simple explanation is that occlusion of the hydrophobic patch of the putative binding epitope prevents nonspecific aggregation of the toxin with itself or other host proteins. Another explanation is that the N-terminal amino acids play a role in the formation of the environmentally stable crystalline inclusions.

The specificity-distinguishing residues are also indicated by comparison of the Cry2Aa structure with the structure of the highly homologous (87% sequence identity) Cry2Ab that is inactive against some Cry2Aa target species (Figure 2a). Chimeric-scanning mutagenesis [24, 25] defines a continuous 106 amino acid block, 307–412, of specificity-distinguishing residues. (Specifically, [25] demonstrated that substitution of residues 278–340 resulted in loss of dipteran-specific activity in Cry2Aa, while [24] demonstrated that substitution of residues 307–382 conferred dipteran-specific activity to Cry2Ab. Thus, in our discussion, we adopt residue 307 as the N-terminal boundary of the specificity-conferring sequence in Cry2Aa.) Within these 106 amino acids, there are 23 residues that differ between Cry2Aa and Cry2Ab (sequence alignment presented in Figure 5). Most of the Cry2Aa-Cry2Ab amino acid differences lie within or about the domain II/III 800 Å² hydrophobic patch (Figure 1b) and surrounding residues from the β5-β6, β7-β8, and β4-β5 loops (Figure 1a). The picture of the putative toxin-receptor binding surface that emerges is that of an 800 Å² hydrophobic region surrounded by three loops, those joining β4-β5, β5-β6, and β7-β8, which are also a part of the putative binding site. The three loops contain hydrophilic side chains that may be involved in specific hydrogen bonding with the receptor and so signal a portion of the site that could be mutated both to probe these interactions and to alter specificity.

The proximity of this surface to solvent-exposed loops of the lectin-like domain III is consistent with the finding that domain III plays a role in the fine tuning susceptibility of different species. This has been demonstrated by replacement of domain III [30, 31] to make chimeric toxins with altered specificity characteristics. The N-ter-

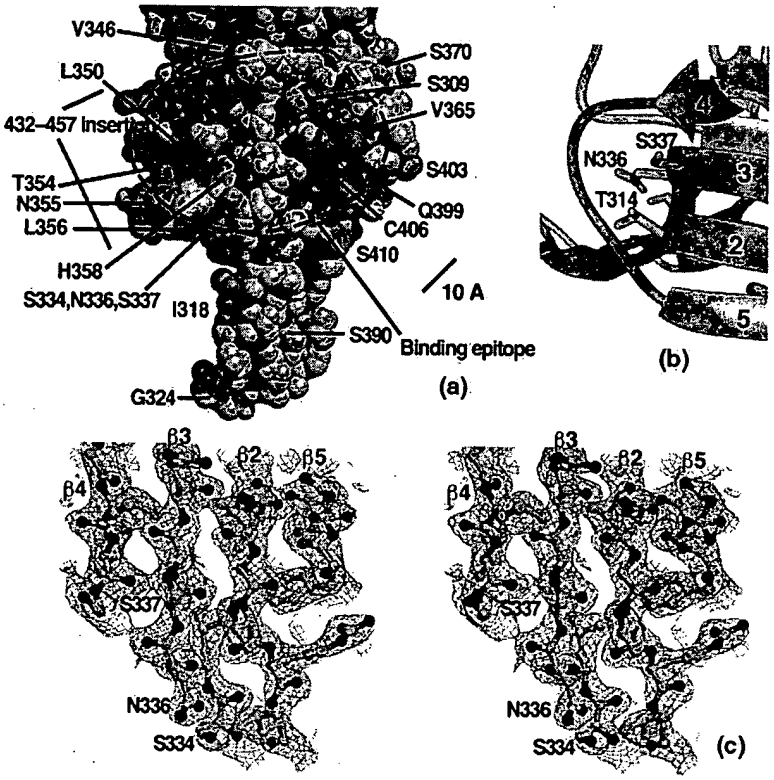


Figure 2. Space-Filling Representation of Cry2Aa Specificity-Conferring Residues, Detail of Buried D Block Residues, and Electron Density Covering Buried D Block Residues

(a) Space-filling model of Cry2Aa domains II and III with the N terminus and membrane-inserting domain I removed. The orientation reflects a −20° rotation relative to that shown in Figure 1a. The results of chimeric-scanning [24, 25] mutagenesis experiments are projected onto the van der Waals surface of Cry2Aa. The residues colored green and cyan are single amino acid differences between Cry2Aa and Cry2Ab in block L (residues 341–412). The residues colored yellow and orange are single amino acid differences between Cry2Aa and Cry2Ab in block D (residues 307–340). The bar represents an approximate 10 Å scale.

(b) Packing of D block residues behind the β4-β5 loop. The β4-β5 loop contains L block specificity determinants with which the buried D block residues interact.

(c) Electron density for the putative receptor binding site covering residues of the β sheet behind the β4-β5 loop.

minal strand β12a of domain III is not present in the three-dimensional structures of Cry1Aa or Cry3Aa. The turn between this strand and the rest of domain III is functionally replaced almost exactly by a loop that connects β3 and β4 of domain II in the homologous Cry1Aa and Cry3Aa structures (Figures 1a and 3, shown in magenta). This functionally conserved β12a motif occupies the same region of the structure as the β3-β4 turn in Cry1Aa and Cry3Aa, so it implies conservation of a functional role in protecting the hydrophobic portion of the putative receptor binding surface implicated by the homolog substitutions.

Chimeric-scanning mutagenesis identifies fairly large

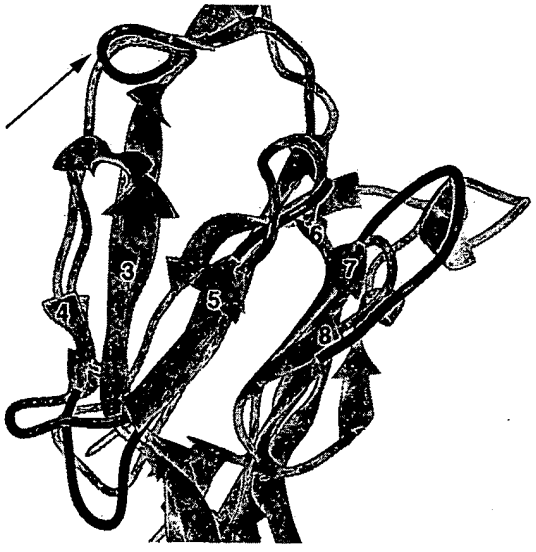


Figure 3. Detail of Ribbon Diagram Overlap of Cry2Aa and Cry1Aa

The Cry1Aa domains have been independently fit to those of Cry2Aa. The functionally important loops delimiting the putative toxin-receptor binding epitope are shown in green (Cry2Aa) and blue (Cry1Aa). The Cry2Aa insertion, relative to Cry3Aa and Cry1Aa, before β12 at the N terminus of domain III is shown in magenta, while the corresponding loop from Cry1Aa is shown in cyan (see arrow).

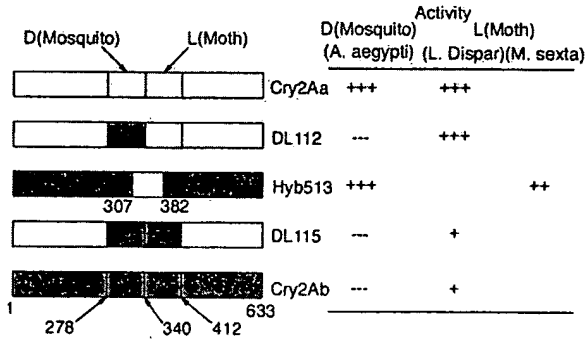


Figure 4. Schematic Representation of Chimeric-Scanning Mutagenesis Data

The first and last bands represent the Cry2Aa and Cry2Ab sequences, respectively. The middle bands represent chimeric combinations in which gray regions correspond to Cry2Ab sequence and white regions correspond to Cry2Aa sequence. For all bands, except that corresponding to Hyb513, the three central vertical bars represent amino acids 278, 340, and 412. For Hyb513, the two central vertical bars represent amino acids 307 and 382. The activity designations represent an approximate log scale. For example, the (+++) activity designation for chimera DL112 corresponds to an ID_{50} of 126 (85.7–187) ng, while the (+) designation for chimera DL115 corresponds to an ID_{50} of 3,200 (1,340–51,900) ng; the confidence intervals correspond to 2σ.

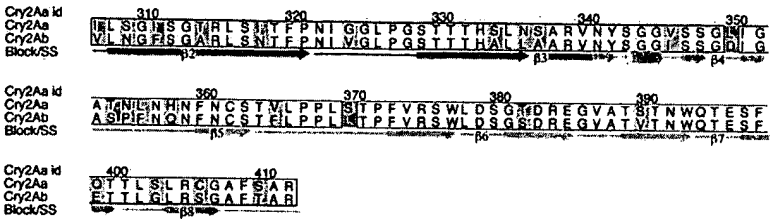


Figure 5. Detail Sequence Alignment of Cry2Aa and Cry2Ab

Sequence alignment of the D and L block regions of Cry2Aa and Cry2Ab generated using ALSCRIPT [51]. In the alignment, identical amino acids are unmarked; similar residues (as defined by ALSCRIPT) are colored yellow, while dissimilar residues are marked green. The secondary structure associated with sequence is presented in the lowermost row. The block of secondary structure associated with D block residues is colored magenta, while that associated with L block residues is colored cyan.

regions of the protein sequence that confer differential specificity to Diptera and Lepidoptera [25] (Figure 4). In Figure 4, the first band represents the sequence of Cry2Aa with its high level of activity (+++) against both Lepidoptera and Diptera. The last band represents the Cry2Ab sequence that exhibits negligible activity (---) against Diptera and up to one order of magnitude lower activity against Lepidoptera when compared with Cry2Aa. The second band (DL112) represents a Cry2Aa chimera that contains the Cry2Ab sequence for the block D residues 307–340 (dipteran-specific). This chimera has negligible activity against Diptera and is suggested to have reduced activity (at the 1σ level) against Lepidoptera, indicating that block D correlates with dipteran specificity. The activity profile of a reverse chimera (the third band) [24], in which Cry2Ab contains the block D sequence from Cry2Aa, shows a more significant reduction than DL112 against Lepidoptera (of a different species) but is only reduced 20-fold toward Diptera versus Cry2Aa. Thus antidipteran activity tracks with the D block of Cry2Aa.

The fourth band (DL115) represents a Cry2Aa chimera that contains the Cry2Ab sequence for the dipteran-disfavoring D block and for a neighboring region of sequence, the lepidopteran-disfavoring L block (residues 341–412). The activity profile of this construct against both Diptera and Lepidoptera most closely parallels that of Cry2Ab, which is consistent with blocks D and L encoding essentially all of the differential specificity determinants. In summary, the differential specificity for Diptera in Cry2Aa depends on block D, while that for Lepidoptera depends on block L. Maximal activity

Table 2. Solvent Accessible Surface Areas, Contacts within 3.4 Å, and Hydrogen Bonds for the Specificity-Conferring Residues in Cry2Aa

Residue	Exposed Surface (Å²)	Exposed Surface Beyond C_β (Å²)	Contacts
			Dipteran Specificity-Conferring
Ile307	6	4	Ser309,Ser343,Gly481,(Met483),(Tyr342)
Ser309	26	26	Asn341,Ile307,Thr364,(Ser363)
Ile311	1	0	Cys362,(Arg339),(Asn361)
Thr314	7	7	Ser337,Asn357,Asn336,His358,(Asn359)
Ile318	91	89	Thr332,(Thr331)
Gly324	78	0	
Ser334	5	5	Leu316,Asn336,(Phe409),(Gln399),(Arg315)
Asn336	6	6	Thr314,Ser334,Ala460,Ala353,(Gly313),(Ile351)
Ser337	0	0	Thr314,Ala353,His358
			Lepidopteran Specificity-Conferring
Val346	39	34	Tyr342,(Asn303),(Gly344)
Leu350	27	26	Asn449,Ile450
Thr354	50	26	Glu451
Asn355	109	76	(Pro457)
Leu356	60	43	(Ala353)
His358	43	14	Ser312,Thr314,Ser337,(Gly313)
Val365	107	75	(Asn336)
Ser370	68	39	Pro367,(His21)
Thr382	54	24	(Asn392),(Thr391)
Ser390	9	9	Ser329,Thr331,Asp383
Gln399	33	33	Val374,Arg375,Arg405,(Leu404)
Ser403	93	73	
Cys406	37	27	Ser397,Phe398,Cys362
Ser410	89	72	

All data were calculated for the activated toxin using HBPLUS [50]. Entries in the left-most column are the 23 specificity-conferring residues. Entries in the right-most column conform to hydrogen bonding geometry, except for those enclosed in parentheses that are van der Waals contacts. Bold entries in the right-most column identify specificity-conferring residues also found in the left-most column.

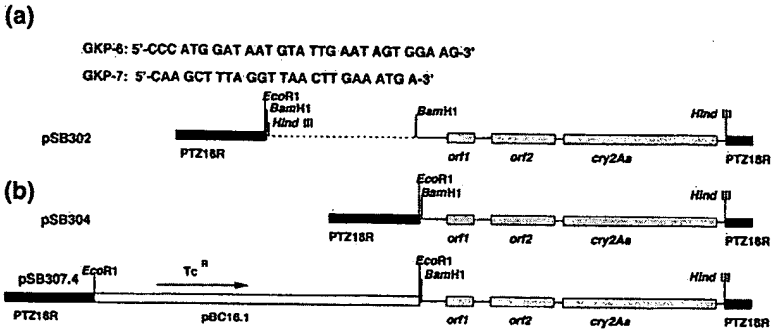


Figure 6. Restriction Maps Detailing the Construction of Plasmid pSB307

(a) Nucleotide sequences of the oligonucleotides GKP-6 and GKP-7.

(b) Restriction maps of pSB302, pSB304, and pSB307.

against Lepidoptera, as seen in Cry2Aa, still requires some contribution from block D (sequence alignment presented in Figure 5).

Figure 2a projects the Cry2Aa/Cry2Ab homolog differences onto the van der Waals surface of Cry2Aa (for clarity, only domains II and III are shown). In the D block, there are nine residues that differ between Cry2Aa and Cry2Ab. Surprisingly, most of these are buried. The notable exceptions are Ile318 and Gly324 (Asn and Val, respectively, in Cry2Ab), which are distant from the putative binding epitope, and the moderately exposed Ser309 (Asn in Cry2Ab) within the putative binding epitope (Table 2). Ile307 and Ile311 are found packed behind exposed residues on the putative binding surface. Almost half of the variant residues from block D (Thr314, Ser334, Asn336, and Ser337) are in a cluster that is packed behind the β4-β5 loop presented from within the 72-residue L block (Figures 2b and 2c).

Two of these buried variant residues, Thr314 and Ser337, make side chain-main chain hydrogen bonds with the β4-β5 loop. A third residue, Asn336, makes main chain-main chain hydrogen bonds with the β4-β5 loop, and Thr314 makes side chain-side chain hydrogen bonds with Ser334. In the less active homolog, Cry2Ab, these residues are replaced with approximately isosteric nonhydrogen bonding residues, suggesting that this pattern of substitutions abolishes affinity for the dipteran receptor (Thr314Ala, Ser334Ala, Asn336Leu, and Ser337Ala). It is conceivable that the Ile318Val and Gly324Val substitutions are part of a region of the protein that interacts only with the receptor(s) found in dipteran species and shares some components with the putative binding epitope that we identify. However, we speculate that the same exposed surface area binds to the lepidopteran and dipteran receptors. In this model, these solvent inaccessible residues behind the putative receptor binding surface may serve to alter the conformation of the β4-β5 loop, with its several hydrophilic specificity-determining residues. Similar modulation of specificity in protein-protein interactions by noncontact residues is seen in the context of immunoglobin residues that affect conformation of the complementarity-determining residues (CDR) at the binding surface [34]. Likewise, affinity maturation of a Fab/antigen complex results in the optimization of antibody/antigen binding by residues 15 Å from the interaction surface [35].

The structures of Bt toxins provide a template for design and discovery of changes that alter receptor targeting in order to either broaden selectivity for better field efficacy, prolong the life of existing agents, or avoid unwanted effects on nontarget organisms. Resistance to Bt toxins is recognized as a potential limitation in their application. Early studies concluded that recessive genes controlled the inheritance of Bt resistance. However, a recent study suggests that Bt resistance can be inherited as an incompletely dominant autosomal gene [36]. The authors note that such a mechanism of Bt resistance inheritance in the field would significantly reduce the usefulness of the high dose/refuge strategy of resistance management in which some mates are not challenged with toxin. Knowledge of any presumed modifications in the receptor that cause resistance can potentially instruct rational protein engineering of the receptor binding surface to yield toxins that might bypass resistance and still bind to the modified receptor of resistant insect species.

Potential collateral effects upon nontarget insect species [36] and effects upon nontarget predatory insects that consume target insect species [37] have been attributed to Bt GM crops. The structures provide a blueprint for focused mutagenesis followed by screening to select for each specific target species in a particular crop, so as to diminish collateral toxicity to nontarget species. By shedding light on the molecular basis of toxin-host receptor recognition, the structure provides a foundation for engineering Bt-based toxin genes that may develop broader insect species specificity, species selectivity tuned to reduce collateral impact upon nontarget species, and longer field efficacy.

Biological Implications

We have determined the three-dimensional structure of the insecticidal toxin Cry2Aa in order to understand the structural determinants of toxin specificity. Genetically modified (GM) crops that express insecticidal protein toxins are an integral part of modern agriculture. Proteins normally produced by different strains of *Bacillus thuringiensis* (Bt) during sporulation mediate a species-specific pathogenicity of Bt toward insect larvae of the target species and are the active agents in the majority of biorational pesticides and insect-resistant transgenic crops. Though promising as a crop protection alternative, problems exist with transgenic crops. Bt GM crops may pose a threat to nontarget insect species [16] or to nontarget predatory insects that consume target insect species [37]. In addition, resistance to Bt toxins is recognized as a potential limitation to their application that is ecologically friendlier than traditional organic pesticides. For instance, EPA approval of Bt GM maize was

contingent upon the establishment of viable resistance management strategies [36].

Cry2Aa is among an unusual subset of crystalline (Cry) proteins possessing broad insect species specificity by exhibiting high specific activity against larvae from two insect orders, Lepidoptera and Diptera [24, 25], of agricultural and public health significance. Also, the Cry2Aa protoxin is significantly smaller (72 kDa) than those of the Cry1 proteins (~135 kDa) in the current generation of transgenic crops. Since gene size can be a limiting factor of protein expression in plants, transgenic constructs based upon Cry1 usually express a smaller portion of the gene that contains essentially the activated toxin. Cry protoxins are presumed to be more environmentally stable than the activated toxins; hence, transgenic constructs that express the Cry2Aa protoxin could deliver higher toxin doses in the field due to greater stability [22]. Also, expression of the protoxin reduces collateral damage to nontarget insect species since it depends on specificity of the host proteases for activation [3, 37]. Chloroplast-directed overexpression of the Cry2Aa protoxin has been demonstrated and shows expression levels equivalent to 2%–3% total soluble protein in transformed leaves [22]. Such high levels of expression, 20- to 30-fold higher than current nuclear transgenics, could diminish the opportunity for developing resistance by significantly increasing toxin dose at the initial encounter with the insect.

Cry2Ab, an 87% sequence identical homolog of Cry2Aa, has negligible activity against dipteran species and 3- to 8-fold less activity against certain lepidopteran species [25, 26]. Also, there exists a unique body of chimeric-scanning mutagenesis data in the Cry2Aa/Cry2Ab system that has identified determinants of species specificity in the amino acid sequence [24, 25]. Correlating the structure with chimeric-scanning data indicates that the putative receptor binding epitope of Cry2Aa lies on the core β sheet and differs from the end of the β sheet apical loops of domain II, as suggested from structures of the other Cry toxins [28, 29]. Thus, a target surface is defined for directed mutagenesis that may focus engineering of the toxin either to develop broader insect species specificity, species selectivity tuned to reduce collateral impact upon nontarget species, or longer field efficacy. Until now, the search for new insecticidal bacterial toxins involved collection and assay of novel isolates of Bt and other bacteria known to have insecticidal activity. Recent reports describe the isolation of bacterial species that produce new classes of insecticidal toxin [38]. These structure data may permit rational engineering of insecticidal Cry toxins with desired characteristics.

Experimental Procedures

Cloning of Cry2Aa

Oligonucleotide primers flanking the coding region of *cry2Aa* were generated based on the published sequence of the gene from Bt *kurstaki* strain HD-1 [26]. Primer GKP-6 is a 29-mer that corresponds to the N-terminal 26 nucleotides of the coding region (Figure 6a). Primer GKP-7 is a 25-mer that corresponds to a fragment overlapping the HindIII site that is located ~350 nucleotides downstream from the stop codon (Figure 6a). Plasmid DNA isolated from Bt *kurstaki* HD-1 served as a template for the PCR reaction. The resulting 2100 bp fragment was purified and served as the probe used to identify the Cry2Aa operon with its accompanying open reading frames. The hexamer-primed labeling method was used to incorporate ^{32}P-dCTP into the probe.

Previously, it was indicated that the entire gene, including the coding region and the promoter, is present on a 5.0 kb HindIII fragment [26] of a plasmid isolated from Bt *kurstaki* HD-1. The 3.5–7 kb fragments obtained by HindIII digestion of plasmid DNA isolated from Bt *kurstaki* HD-1 were ligated into an *E. coli* cloning vector, pTZ18R (Pharmacia, vecbase accession #VB0071) and were used to transform *E. coli* DH5 cells by electroporation. Electroporated DH5 cells were plated onto LB-Amp^{50} plates containing X-gal and IPTG for color selection. The presence of the *cry2Aa* gene in the transformed colonies (white) was confirmed by hybridization of the PCR-generated probe. Restriction analysis was used to confirm that the clones contained inserts with the *cry2Aa* gene and also to establish the orientation with which the fragment was inserted into pTZ18R. The results of this analysis revealed that one of the clones corresponded to the orientation designated pSB302 (Figure 6b), while two clones had the opposite orientation and were designated pSB303. pSB304, obtained by deleting the 1.2 kb-BamHI fragment (dotted line in Figure 6b), was also transformed into DH5.

Total protein analysis for proteins produced by *E. coli* strain DH5 carrying pSB302, pSB303, or pSB304 was performed by SDS-PAGE. A protein band of molecular weight 62 kDa, absent in the original DH5 cells, was observed in all of the clones examined. The level of expression was the highest in those cells bearing pSB304. Most of the toxin could be found in the pelletable fraction following sonication of the cells. Samples were evaluated for biological activity by bioassay using *Manduca sexta* as the target insect. All of the clones (pSB302, pSB303, and PSB304) were active with LD_{50} values of ~500 ppm.

The pSB304 plasmid retains a unique EcoR1 site, ~200 nucleotides upstream of the *cry2Aa* promoter, into which the EcoR1-linearized *Bacillus cereus* vector pBC16.1 (GenBank accession number U32369) was cloned (Figure 6b). The resulting clone was used to transform *E. coli* DH5, and clones containing the new plasmid were designated pSB307. Confirmation of the identity of the new plasmid and determination of the orientation of the pBC16.1 insert, with respect to the *cry2Aa* gene, was made by restriction mapping. One of the plasmids, pSB307.4, was transformed into Bt *cryB* (a cry$^-$ strain) by electroporation. The plasmid content of these isolates was verified by restriction mapping.

Cry2Aa expressed well in Bt cryB cells transformed with pSB307.4, and the protein formed crystalline (rhombohedral) inclusions. The cells were harvested by centrifugation, washed with water, and lyophilized. Dried cell mass was added to the insect diet and fed to *M. sexta* larvae. The results confirmed that Bt *cryB* (pSB307.4) exhibited high insecticidal activity.

Protein Expression and Purification

The plasmid (pSB307.4) containing the Cry2Aa operon, with its accompanying open reading frames, was used to transform the cry$^-$ strain of Bt (*cryB*) as previously described [39]. Cry2Aa was purified from the crystalline inclusions produced in the cells. Inclusions were harvested by cell lysis and centrifugation. Crystalline inclusions were washed repeatedly with 0.5 N NaCl to remove proteases and were transferred to buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]) with 2% mercaptoethanol. Titrating the pH to 10.5, using NH_4OH, solubilized the protein from the crystalline inclusion bodies. The protein was purified by Sephacryl S300HR column chromatography as described [40] and concentrated by ultrafiltration to 10 mg ml^{-1}.

Crystallization and Structure Determination

For recrystallization, hanging drops of the resulting concentrated protein (10 μl concentrated protein buffered as described above) were equilibrated against wells that contained Tris buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]). Crystallization was induced by the gradual shift to neutral pH as the mobile NH_3 diffused from the drops. Crystals were transferred to storage buffer (50 mM PIPES, 250 mM NaCl [pH 6.8]) with 2% mercaptoethanol. The resulting crystals are in spacegroup $P4_32_12$; unit cell constants a = 85.6 Å, c = 163.9 Å. They have one monomer in the asymmetric unit, an

estimated 34% solvent content, and diffract to ~3.0 Å using Cu K_α X-rays from a rotating anode generator and to 2.0 Å at a synchrotron source after flash freezing.

For the collection of data at 100K, the crystals were transferred in three steps to a final 20% solution of cryo-protectant (a 1:1 mixture of 1,2-propane diol and glycerol) and storage buffer and flash frozen in a cold nitrogen stream. X-ray diffraction data were collected at SSRL beamline 7.1 using a wavelength of 1.08 Å. Intensity data were integrated, scaled, and merged using HKL [41]. The overall Wilson B factor (3.0 Å < d < 2.2 Å) was 14 Å^2.

De novo phasing was achieved using multiple isomorphous replacement after attempts to find a molecular replacement solution to the phase problem employing the available coordinates of Cry3Aa and Cry1Aa were unsuccessful. The heavy atom derivatives (Table 1) were solved from difference Patterson maps as displayed using XtalView [42]. Difference Fourier inspection for minor sites and refinement of the heavy atom positions, occupancies, and B factors was completed in PHASES [43]. The resulting protein electron density map was subjected to solvent-flipping density modification, as implemented in Solomon [44]. The helical bundle was apparent in 5 Å maps; at 3 Å resolution, the correct enantiomorph was clear from its stereochemistry. Using Cry1Aa as the initial building template, polyalanine versions of the helical and jellyroll domains were manually positioned using O [45], and the fit was optimized using the real-space refinement package ESSENS [46]. Positional and simulated annealing refinement were carried out using the maximum likelihood target of XPLOR 3.85× [47].

Acknowledgments

We thank D.H. Dean, E.A. Zhukovsky, J. Finer-Moore, and R.J. Fletterick for helpful discussions during the course of this investigation. We thank V. Ramalingam for assistance in crystallization and data collection and G.K. Powell for providing us with the Cry2Aa clone. This work is based upon research conducted at SSRL, which is funded by the Department of Energy, Office of Basic Energy Sciences. This work was supported by the National Institutes of Health (GM-244485 to R.M.S).

Received: December 27, 2000
Revised: April 4, 2001
Accepted: April 6, 2001

References

1. Schnepf, E., et al., and Dean, D.H. (1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. *62*, 775–806.
2. Tojo, A., and Aizawa, K. (1983). Dissolution and degradation of *Bacillus thuringiensis* δ-endotoxin by gut juice protease of the silkworm *Bombyx mori*. Appl. Environ. Microbiol *45*, 576–580.
3. Aronson, A.I., Han, E.S., McGaughey, W., and Johnson, D. (1991). The solubility of inclusion proteins from *Bacillus thuringiensis* is dependent upon protoxin composition and is a factor in toxicity to insects. Appl. Environ. Microbiol *57*, 981–986.
4. Choma, T., Surewicz, W.R., Carey, P.R., Pozsgay, M., Raynor, T., and Kaplan, H. (1990). Unusual proteolysis of the protoxin and toxin from *Bacillus thuringiensis*. Structural implications. Eur. J. Biochem. *189*, 523–527.
5. Hofmann, C., Vanderbruggen, H., Höfte, H., Van Rie, J., Jansens, S., and Van Mellaert, H. (1988). Specificity of *Bacillus thuringiensis* δ-endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target insect midguts. Proc. Nat. Acad. Sci. USA *85*, 7844–7848.
6. Wolfersburger, M.G., Hofmann, C., and Lüthy, P. (1986). Interaction of *Bacillus thuringiensis* with membrane vesicles isolated from lepidopteran larval midgut. Zbl. Bakt. Mikrobiol. Hyg. I. Suppl. *15*, 237–238.
7. Van Rie, J., Jansens, S., Höfte, H., Degheele, D., and Van Mellaert, H. (1989). Specificity of *Bacillus thuringiensis* δ-endotoxins: importance of specific receptors on the brush border membrane of the mid-gut of target insects. Eur. J. Biochem. *186*, 239–247.
8. Slatin, S.L., Abrams, C.K., and English, L. (1990). Delta-endotoxins form cation-selective channels in planar lipid bilayers. Biochem. Biophys. Res. Commun. *169*, 765–772.
9. English, L., Readdy, T.L., and Bastian, A.E. (1991). Delta-endotoxin-induced leakage of $^{86}Rb^+$-K^+ and H_20 from phospholipid vesicles is catalyzed by reconstituted midgut membrane. Insect Biochem. *21*, 177–184.
10. Schwartz, J.L., Garneau, L., Savaria, D., Masson, L., and Brousseau, R. (1993). Lepidopteran-specific crystal toxins from *Bacillus thuringiensis* form cation- and anion-selective channels in planar lipid bilayers. J. Membrane Biol. *132*, 53–62.
11. Harvey, W.R., and Wolfersberger, M.G. (1979). Mechanism of inhibition of active potassium transport in isolated midgut of *Manduca sexta* by *Bacillus thuringiensis* endotoxin. J. Exp. Biol. *83*, 293–304.
12. Harvey, W.R., Cioffi, M., and Wolfersberger, M.G. (1986). Transport physiology of lepidopteran midgut in relation to the action of *B.t.* delta-endotoxin. In Fundamental and Applied Aspects of Invertebrate Pathology, J.M. Vlak, D. Peters, and R.A. Samson, eds. (Wageningen, The Netherlands: Grafisch Berdrijf Ponsen and Looijen), pp. 11–14.
13. Knowles, B.H., and Ellar, D.J. (1987). Colloid-osmotic lysis is a general feature of the mechanism of *Bacillus thuringiensis* d-endotoxins with different insect specificity. Biochem. Biophys. Acta *924*, 509–518.
14. Wolfersberger, M.G. (1992). V-ATPase energized epithelia and biological insect control. J. Exp. Biol. *172*, 377–386.
15. Crickmore, N., et al., and Dean, D.H. (1998). Revision of the nomenclature of the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. *62*, 807–813.
16. Losey, J.E., Rayor, L.S., and Carter, M.E. (1999). Transgenic pollen harms monarch larvae. Nature *399*, 214.
17. Van Rie, J., McGaughey, W.H., Johnson, D.E., Barnett, B.D., and Van Mellaert, H. (1990). Mechanism of insect resistance to the microbial insecticide *Bacillus thuringiensis*. Science *247*, 72–74.
18. McGaughey, W.H., Gould, F., and Gelertner, W. (1998). Bt resistance management. Nature Biotechnol. *16*, 144–146.
19. Yamamoto, T., and McLaughlin, R.E. (1981). Isolation of a protein from the parasporal crystal of *Bacillus thuringiensis* var. *kurstaki* toxic to the mosquito larva *Aedes taeniarhynehus*. Biochem. Biophys. Res. Commun. *103*, 414–421.
20. Donovan, W.P., Dankocsik, C.C., Gilbert, M.P., Gawron-Burke, M.C., Groat, R.G., and Carlton, B.C. (1988). Amino acid sequence and entomocidal activity of the P2 crystal protein. An insect toxin from *Bacillus thuringiensis* var. *kurstaki*. J. Biol. Chem. *263*, 561–567.
21. Yamamoto, T., and Powell, G.K. (1993). *Bacillus thuringiensis* crystal proteins: recent advances in understanding its insecticidal activity. In Advanced Engineered Pesticides. L. Kim, ed. (New York: Marcel Dekker), pp. 3–42.
22. Kota, M., Daniell, H., Varma, S., Garczynski, S.F., Gould, F., and Moar, W.J. (1999). Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. Proc. Nat. Acad. Sci. USA *96*, 1840–1845.
23. English, L., et al., and Slatin, S.L. (1994). Mode of action of CryIIA: a *Bacillus thuringiensis* delta-endotoxin. Insect Biochem. Mol. Biol. *24*, 1025–1035.
24. Widner, W.R., and Whiteley, H.R. (1990). Location of the dipteran specificity region in a lepidopteran-dipteran crystal protein from *Bacillus thuringiensis*. J. Bact. *172*, 2826–2832.
25. Liang, Y., and Dean, D.H. (1994). Location of a lepidopteran specificity region in insecticidal crystal protein CryIIA from *Bacillus thuringiensis*. Mol. Microbiol. *13*, 569–575.
26. Widner, W.R., and Whiteley, H.R. (1989). Two highly related insecticidal crystal proteins of *Bacillus thuringiensis* subsp. *kurstaki* possess different host range specificities. J. Bacteriol. *171*, 965–974.
27. Audtho, M., Valaitis, A.P., Alzate, O., and Dean, D.H. (1999). Production of chymotrypsin-resistant *Bacillus thuringiensis* Cry2Aa1 delta-endotoxin by protein engineering. Appl. Environ. Microbiol. *65*, 4601–4605.
28. Li, J., Carroll, J., and Ellar, D.J. (1991). Crystal structure of insec-

ticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. Nature *353*, 815–821.
29. Grochulski, P., et al, and Cygler, M. (1995). *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J. Mol. Biol. *254*, 447–464.
30. Lee, M.K., Young, B.A., and Dean, D.H. (1995). Domain III exchanges of *Bacillus thuringiensis* Cry1A toxins affect binding to different gypsy moth midgut receptors. Biochem. Biophys. Res. Commun. *216*, 306–312.
31. de Maagd, R.A., et al., and Bosch, D. (1996). Domain III substitution in *Bacillus thuringiensis* delta-endotoxin CryIA(b) results in superior toxicity for *Spodoptera exigua* and altered membrane protein recognition. Appl. Environ. Microbiol. *62*, 1537–1543.
32. Schwartz, J.L., Potvin, L., Chen, X.J., Brousseau, R., Laprade, R., and Dean, D.H. (1997). Single-site mutations in the conserved alternating-arginine region affect ionic channels formed by CryIAa, a *Bacillus thuringiensis* toxin. Appl. Environ. Microbiol. *63*, 3978–3984.
33. Von Tersch, M.A., Slatin, S.L., Kulesza, C.A., and English, L.H. (1994). Membrane-permeabilizing activities of *Bacillus thuringiensis* coleopteran-active toxin CryIIIB2 and CryIIIB2 domain I peptide. Appl. Environ. Microbiol. *60*, 3711–3717.
34. Foote, J., and Winter, G. (1991). Antibody framework residues affecting the conformation of the hypervariable loops. J. Mol. Biol. *224*, 487–499.
35. Wedemayer, G.J., Patten, P.A., Wang, L.H., Schultz, P.G., and Stevens, R.C. (1997). Structural insights into the evolution of an antibody combining site. Science *276*, 1665–1669.
36. Huang, F., Buschman, L.L., Higgins, R.A., and McGaughey, W.H. (1999). Inheritance of resistance to *Bacillus thuringiensis* toxin (Dipel ES) in the European corn borer. Science *284*, 965–967.
37. Hilbeck, A., Moar, W.J., Pusztai-Carey, M., Filippini, A., and Bigler, F. (1999). Prey-mediated effects of Cry1Ab toxin and protoxin and Cry2A protoxin on the predator *Chrysoperla carnea*. Entomol. Exp. Appl. *91*, 305–316.
38. Bowen D., et al., and ffrench-Constant, R.H. (1998). Insecticidal toxins from the bacterium *Photorhabdus luminescens*. Science *280*, 2129–2132.
39. Sasaki, J., et al., and Yamamoto, T. (1996). Insecticidal activity of the protein encoded by the cryV gene of *Bacillus thuringiensis kurstaki* INA-02. Curr. Microbiol. *32*, 195–200.
40. Yamamoto, T. (1989). Identification of entomocidal toxins of *Bacillus thuringiensis* by high-performance liquid chromatography. ACS Symposium Series, *432*, 46–60.
41. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
42. McRee, D. (1993). Practical Protein Crystallography. (San Diego, CA: Academic Press).
43. Furey, W., and Swaminathan, S. (1997). PHASES-95: A program package for the processing and analysis of diffraction data from macromolecules. In Methods Enzymology C. Carter and R. Sweet, eds. (Orlando FL: Academic Press), pp. 307–326.
44. Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F-1 Atpase. Acta Crystallogr. D *52*, 30–42.
45. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
46. Kleywegt, G.J., and Jones, T.A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. Acta Crystallogr. D *53*, 179–185.
47. Brünger, A.T. (1993). X-PLOR Version 3.1 a System for X-ray Crystallography and NMR. (New Haven, CT: Yale University Press).
48. Ferrin, T.E., Huang, C.C., Jarvis, L.E., and Langridge, R. (1988). The MIDAS display system. J. Mol. Graph. *6*, 13–27, 36–37.
49. Nicholls, A. (1992). GRASP Manual. (New York: Columbia University).
50. McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.
51. Barton, G.J. (1993). ALSCRIPT: A tool to format multiple sequence alignments. Protein Eng. *6*, 37–40.

Accession Numbers

The coordinates and structure factors for Cry2Aa have been deposited with the Protein Data Bank (accession code 1I5P).